AFRL-VS-HA-TR-98-0050

TESTING SPECIFICATION MODEL AND REAL-TIME DATA COUPLING SCHEMES AND DEVELOPMENT OF A PROTOTYPE EXECUTIVE SYSTEM FOR SPACE WEATHER PREDICTION

R. W. Schunk
J. J. Sojka
J. V. Eccles
D. C. Thompson
R. L. McPherron

Space Environment Corporation
399 North Main, Suite 325
Logan, UT 84321

20 April 1998

Final Report
21 April 1994 to 20 January 1998

Approved for public release; distribution unlimited

AIR FORCE RESEARCH LABORATORY
Space Vehicles Directorate
29 Randolph Road
AIR FORCE MATERIEL COMMAND
HANSCOM AFB, MA 01731-3010

This technical report has been reviewed and is approved for publication.

IRVING MICHAEL, Contract Manager

GREGORY GINET, Branch Chief

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*

2. REPORT DATE
20 April 1998

3. REPORT TYPE AND DATES COVERED
FINAL REPORT; 4/21/94 to 1/20/98

4. TITLE AND SUBTITLE
Testing Specification Model and Real-Time Data Coupling Schemes and Development of a Prototype Executive System for Space Weather Prediction.

5. FUNDING NUMBERS
Contract
F19628-94-C-0021
PE 63707F
PR 4026 TA GC
WU MA

6. AUTHOR(S)
R. W. Schunk, J. J. Sojka, J. V. Eccles,
D. C. Thompson, R. L. McPherron

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Space Environment Corporation
399 North Main, Suite 325
Logan, Utah 84321

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Air Force Research Laboratory
29 Randolph Road
Hanscom AFB, MA 01731-3010

Contract Manager: Irving Michael/VSBS

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
AFRL-VS-HA-TR-98-0050

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

The main objective of the effort was to develop a multi-level executive system to couple the Air Force specification models and data streams in order to improve space weather predictions. The system was to be based on the physical processes that determine the coupling of adjacent regions of space embodied in the models. A step-by-step procedure was to be employed to develop a workable system. First, the scientific issues concerning a number of possible model-model and model-data coupling schemes were to be studied. Coupling on different levels was to be considered, with the different levels depending on the data available, the quality of the data, the geophysical conditions, and the parameters desired. However, the effort was descoped and redirected and, as a consequence, only part of the work was completed. The completed work involved the construction of an initial ISEM executive system, an analysis of the Kp index, the development of an algorithm that calculates a real-time Dst index, an analysis of the plasma convection and particle precipitation patterns, and the issues concerning data quality.

14. SUBJECT TERMS
Executive System, Specification and Forecast Models, Data Ingestion, Space Weather Prediction

15. NUMBER OF PAGES
80

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	SAR

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std Z39-18
298-102

Table of Contents

Executive Summary....v

1. Introduction....1

2. Executive System....4

3. Kp Index....7

4. Dst Index....9

4.1 Delivery and Implementation at RADEX Corporation
4.2 Specifics of Implementation and Testing in DST.COM Environment
4.3 Assessing the Dst Approximation using Real-Time Data
4.4 Data Problems
4.5 Future Requirements for Quality Dst Approximation
4.6 Conclusion

5. Convection and Precipitation Inputs....12

5.1 Precipitation Pattern
5.2 Convection Pattern

6. Quality Management and Verification....21

6.1 Magnetometer Data
6.2 DMSP J/4 Particle Data
6.3 DMSP IES Data
6.4 DMSP E-Field Data
6.5 Ionospheric Measuring System Data
6.6 DISS Data
6.7 Quality Management and Verification Approach

Appendix....33

Illustrations

1. Initial Version of an ISEM System 5

2. Comparison of the MSM Results Using the Gottingen Kp and Synthetic Kp* for Cases When the Synthetic Kp* Differs from Kp via Standard Deviations of 0.25, 0.50, 0.75, 1.0, and 1.25 8

3. Comparison of the Precipitating Electron Energy Flux Obtained From the Hardy Statistical Model with DMSP Satellite Measurements for Cases When the Agreement is Excellent 13

4. Same as Figure 3, but for Cases When There Are Substantial Differences Between the Hardy Statistical Model and Measurements 14

5. Precipitating Energy Flux Obtained From the Hardy Statistical Model and the Measurements 16

6. Electron Densities at 110 km Calculated with the Hardy Statistical Model and the Measured Precipitation 17

7. Same as Figure 6, Except That the Comparison is for N_mF_2 18

8. Comparison of Horizontal Plasma Drifts Obtained from the Heppner-Maynard Convection Model With Drifts Measured by the DMSP Satellites. 19

9. Same as Figure 8, Except for Active Magnetic Conditions 20

EXECUTIVE SUMMARY

The main objective of the effort was to develop a multi-level executive system to couple the Air Force specification models and data streams in order to improve space weather predictions. The system was to be based on the physical processes that determine the coupling of adjacent regions of space embodied in the models. The system had to be carefully configured because error propagation in coupled models can lead to completely erroneous results under certain conditions. Likewise, when models and data are coupled, inaccurate or incomplete data can result in less reliable predictions than might otherwise be obtained. Therefore, a step-by-step procedure was to be employed to develop a workable system. First, the scientific issues concerning a number of possible model-model and model-data coupling schemes were to be studied. We were supposed to develop the scientific and conceptual basis for implementation of the model coupling. Coupling on different levels was to be considered, with the different levels depending on the data available, the quality of the data, the geophysical conditions, and the parameters desired. However, the effort was descoped and redirected and, as a consequence, only part of the work was completed. The completed work involved the construction of an initial ISEM executive system, an analysis of the Kp index, the development of an algorithm that calculates a real-time Dst index, an analysis of the plasma convection and particle precipitation patterns, and the issues concerning data quality.

1. INTRODUCTION

The main objective of the research was the development of a sophisticated, multi-level executive system to couple the Air Force specification models and data streams in an effort to improve space weather predictions. An executive system had the potential for providing more reliable predictions of space weather than would be obtained if the models were accessed independently. However, when models are coupled, error propagation can lead to completely erroneous results under certain conditions. Likewise, when models and data streams are coupled, inaccurate of incomplete data can result in less reliable predictions than might otherwise be obtained. Because of the uncertainties associated with the possible model-model and model-data coupling schemes, we were supposed to test the coupling schemes in an effort to elucidate potential problems or limitations. Coupling on different levels was to be considered, with the different levels depending on the data available, the geophysical conditions, and the parameters desired. We also were supposed to test and/or develop certain neural networks predictors for forecasting. The end result of our work was to be an advanced scientific software package that would correspond to a workable and well-tested executive system.

The specific tasks that we were originally funded to do are as follows:

Task 1. Run the Air Force specification and forecast models sequentially, which corresponds to the routine running phase recommended for the executive system. This involves

(a) Standardizing the model inputs & outputs.
(b) Reducing outputs to a minimum.
(c) Verifying the integrity of the model outputs.
(d) Developing algorithms to verify the integrity of both the ground-based and satellite data that will be used as inputs to the models.
(e) Verifying the integrity of the convection & precipitation patterns generated by the MSFM.
(f) Testing the model calling sequence for unexpected problems.
(g) Developing algorithms for testing the magnetic indices used in real-time operations.

Task 2. Tests Involving Real-Time Requests

Level 0 corresponds to the routine running phase, and this involves the following:

(a) Develop algorithms to access the global databases generated by the specification and forecast models in the routine running phase. Determine whether interpolation on a 'coarse' model grid or direct retrieval from a 'fine' grid is more appropriate considering parameter accuracy and computer resources.
(b) Develop algorithms based on magnetic indices so that a quality flag can be attached to model parameters returned at this level of the executive system.

Level 1 corresponds to a database search for a measured parameter at a specified

location and time. If the parameter is not available, the database is again searched for supporting measurements that can be used to drive a new model run. At this level, we would:

(a) Define how close in space and time you need to be to satisfy the request, which will be different for different parameters.
(b) Develop algorithms to handle data gaps or unreliable data streams for the main data sources (DISS, TISS, DMSP, magnetometer).
(c) Develop algorithms to determine the minimum data requirements (in quality & quantity) for the supporting measurements to be useful for a new model run.
(d) Use historical data for new model runs to test the real-time operational procedure. This will be done for all models and all of the important geophysical parameters.

Level 2 involves several schemes for coupling the specification and forecast models. We need to test the proposed coupling schemes.

Task 3. Tests Involving Forecast Requests.

(a) Develop and/or test the neural networks that are needed for forecasting the solar and magnetic indices.
(b) Test the quality of Level-1 VSH, IFM & MSFM forecasts using historical data.
(c) Develop algorithms to test the model coupling schemes proposed for level 2 of the executive system. The coupling schemes will be tested for a wide range of geophysical conditions.

Task 4. Tests involving Post Analysis Requests.

Here, more model-model and model-data coupling schemes are possible because more data should be available. We will develop algorithms to access the data and test the various coupling schemes for all of the important geophysical parameters.

Task 5. Develop scientific software for a prototype executive system.

Task 6. Collect ground-based and satellite data for executive system testing.

Task 7. Detailed outline of the training videos needed to cover the solar wind and interplanetary medium, the magnetosphere, the ionosphere, and the thermosphere.

Shortly after the research was initiated, it became apparent that there would be

insufficient funds to fully fund the contract. Consequently, it was decided that the executive system would be developed at the Air Force Research Laboratory in Bedford, Massachusetts, and we would focus on the contract work that would be most beneficial to the Air Force scientists working on the Executive System (later termed Geospace). From then on, the Contract Monitor directed our work. Finally, on September 30, 1996 work stopped due to a lack of funding. A brief description of the tasks that were worked on up to that date is given in what follows:

Task 1. All items in this task have been worked on and good progress has been made on each. SEC has brought to PL's attention concerns with items (d) and (g), which relate to ground-based and satellite data streams as well as the geomagnetic indices Kp and Dst that are required in real-time.

Task 2, 3, & 4. These tasks involve the three operational modes by which the Space Forecast Center responds to requests, namely the real-time, forecast, and post analysis modes. We conducted investigations and/or assessments of various items connected with these tasks, but only for the level 0 issues. **Levels 1 and 2 were not worked on.** In the course of doing this work, we raised several key issues that needed further attention because of adverse effects they have on the overall operations at the Space Forecast Center. The specific details were discussed at quarterly review meetings and were also given in the quarterly reports sent to the Air Force as part of our reporting requirements. The main area of concern was in obtaining a set of geomagnetic indices (Kp and Dst) that is quality controlled, because almost all of the Air Force models require accurate geomagnetic indices. To a lesser extent, similar concerns were found in other data streams. However, progress has been made in resolving these other data stream issues since we brought them to the Air Force's attention. In view of the importance of the Kp and Dst indices to the Space Forecast Center's operations, we view Task 2, Level 0, items (a) and (b) as work that still needs to be done. This is also true for Task 3, item (a).

Task 5. After an initial executive system was constructed, the Contract Monitor directed us not to work on this task.

Task 6. Most of the data needed for this task were collected and they were used to test some algorithms that we developed.

Task 7. This task was not worked on. This final task was envisaged as a product aimed to educate operators at the Space Forecast Center on how to use the executive system, but as noted above, we were instructed not to develop this system.

Subsequently, additional funding was provided to develop a software program that would calculate a real-time Dst index. This FORTRAN program was developed and delivered to the Air Force Research Laboratory on 20 April 1998.

Additional details concerning some of the work that was done under this contract are given in the sections that follow. The complete details can be found in the reports we presented at the Models Review meetings and in previous Quarterly Reports.

2. EXECUTIVE SYSTEM

Before the cutback in funding and the redirection of our work, we created an initial version of the executive system that would control the coupling of the different Air Force models and data streams. The specification and forecast models that we were to consider are:

- Interplanetary Shock Propagation Model (ISPM)
- Solar Wind Transport Model (SWT)
- Magnetospheric Specification Model (MSM)
- Parametrized Real-Time Ionospheric Specification Model (PRISM)
- Ionospheric Forecast Model (IFM)
- Thermospheric Forecast Model (TFM)
- Ionospheric Scintillation Specification and Prediction System (ISSPS)

The satellite and ground-based systems/data we were to consider are:

- SSIES Data from the DMSP Satellites
- Digital Ionospheric Sounding System (DISS)
- Ionospheric Measuring System (TISS)
- USGS Magnetometer Network
- Solar Observations
- NOAA & LANL Geosynchronous Satellites

The executive system was needed to perform the following tasks: (1) Eliminate duplication. For example, several of the specification and forecast models require the plasma convection pattern as an input, and these models were constructed to provide their own convection patterns. This duplication was to be eliminated and the executive system was to provide a convection pattern that all of the specification and forecast models could use; (2) Determine the adequacy of the required empirical models (Solar Flux, magnetic field, etc.); (3) Verify the quality of the data streams that were to be ingested by the specification and forecast models; (4) Verify the integrity of forecast model outputs; and (5) Verify the integrity of the convection and precipitation patterns that were currently available.

The initial version of the ISEM Executive System that we created is shown in Figure 1. Six of the specification and forecast models were acquired and incorporated in the ISEM Executive System, including the ISPM, SWT, MSM, PRISM, IFM and ISSPS models. Also, several well-known empirical models were incorporated in the ISEM Executive System, including auroral precipitation models (Hardy, Spiro), plasma convection models

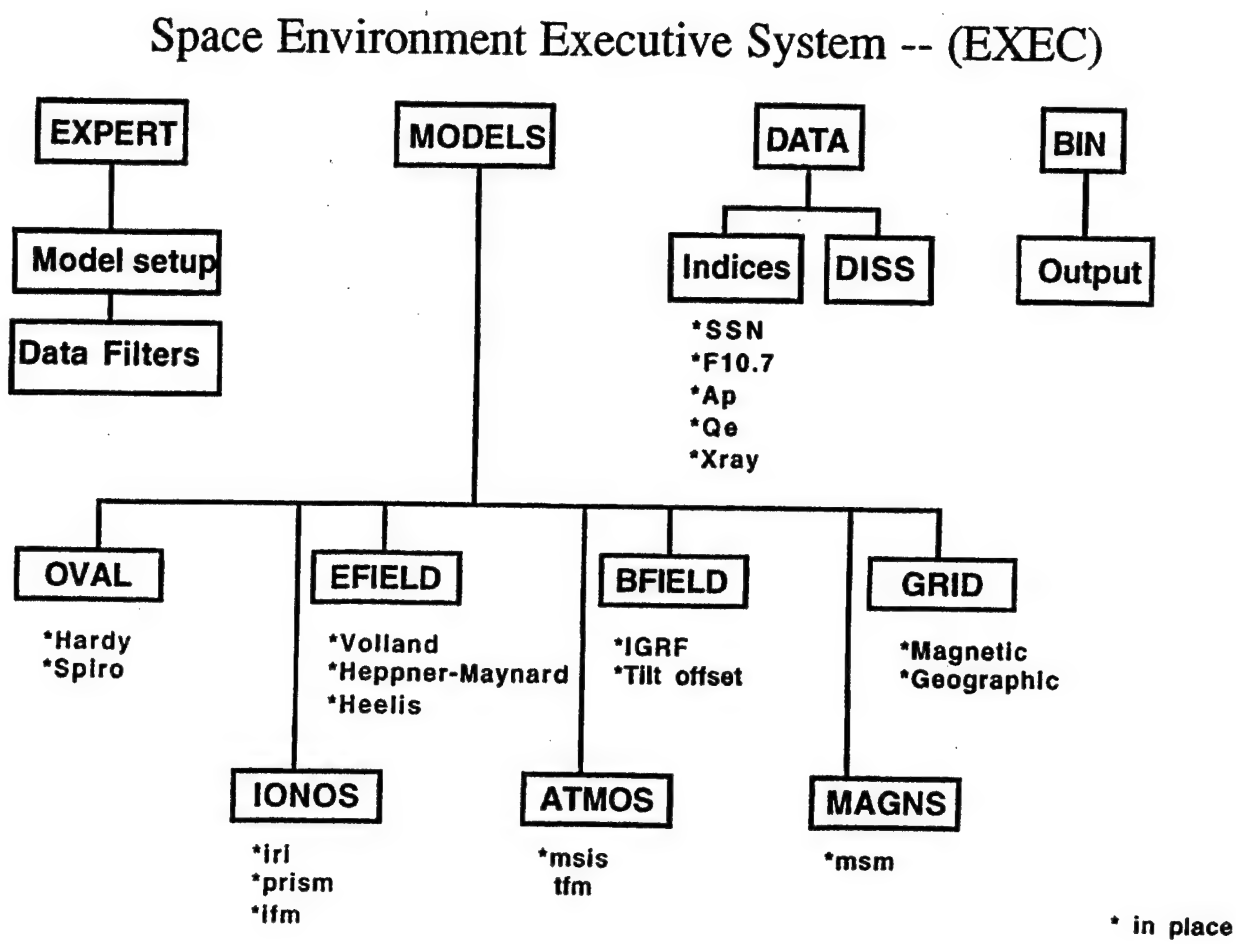


Figure 1. Initial version of an ISEM system.

(Volland, Heelis, Heppner-Maynard), the International Reference Ionosphere (IRI), the MSIS models for neutral densities and winds, and two magnetic field models (IGRF and tilted offset dipole). The system was also configured to ingest geophysical indices in real time (Kp, Ap, F10.7, etc.). Finally, routines for various coordinate transformations were incorporated so that it would be easy to transform data to the different specification and forecast models. At the start of a run, a simple selection procedure allows the operator to select the specification or forecast model to be run and the desired empirical input models. The ISEM Executive System then accesses the required on-line data and runs the model. We were able to successfully test this initial version of an executive system before we were redirected to other tasks.

3. Kp INDEX

The magnetic activity index Kp is an important input for both the Ionospheric Forecast Model (IFM) and the Thermospheric Forecast Model (TFM). However, in 1993, Major Willow Cliffswallow showed that the magnetic activity index (Kp*) used at the Space Forecast Center was not the same as the Standard Göttingen index (Kp). Using data from the one-year period from June 1992 to June 1993 she showed that there were important differences in Kp* and Kp in the range from 0 to about 5. We were subsequently asked to repeat the study for the year 1995 to see if the difference in Kp's persisted. Our results for 1995 were very similar to those obtained by Major Cliffswallow. Table 1 shows the difference between the SFC Kp* and the Göttingen Kp for both the Cliffswallow study and our (SEC) study. The two Kp's were the same 52% of the time in the Cliffswallow study (1992) and 58% of the time in our study (1995). Both studies found the difference distribution shifted to lower Kp, i.e., the average Göttingen Kp was lower than the average SFC Kp*. The source of the difference between Kp determinations probably lies in the SFC (NOAA) algorithm that determines the incremented Kp values. However, our study could not rule out the possibility that the difference is due to a difference in the longitudinal distribution of the magnetometer sites used to calculate Kp.

Table 1. Difference Between Göttingen Kp and SFC Kp*

Gött Kp-SFC Kp*	Cliffswallow study	SEC study
difference = -1	24%	22%
difference = 0	52%	58%
difference = 1	20%	18%

As a follow-up to the above study, we worked with Capt. T. Smith to determine the extent to which the difference between Kp and Kp* affects the magnetospheric Specification Model (MSM) output. Capt. Smith acquired the Göttingen Kp values for a 60-day period and we generated synthetic Kp* values which had standard deviations from Kp of 0.25, 0.50, 0.75, 1.0 and 1.25. The comparison of the MSM results for Kp and Kp* values is given in Figure 2 for the five standard deviation cases. Note that the SFC Kp* index differs statistically from the Göttingen Kp by a standard deviation of 0.75.

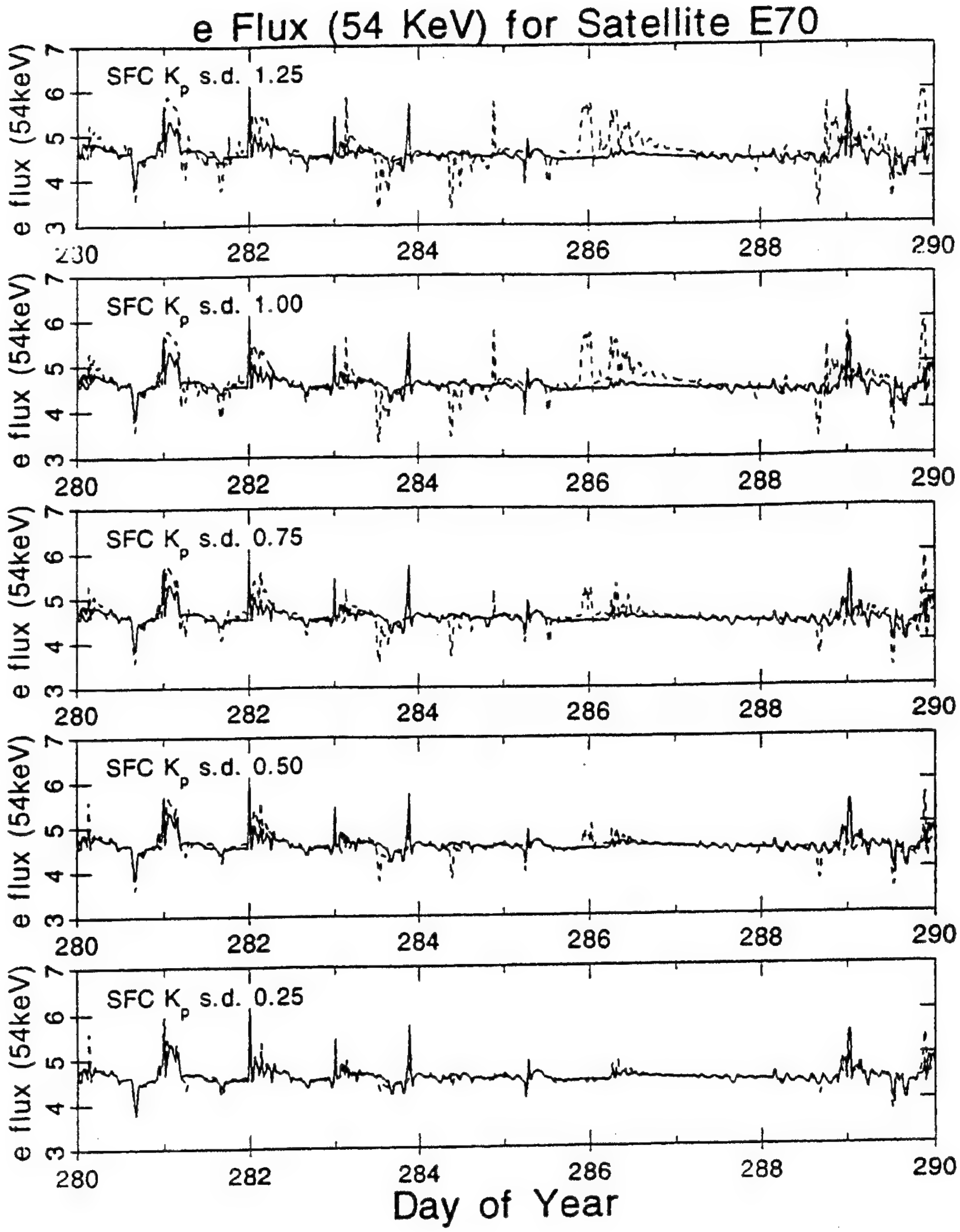


Figure 2. Comparison of the MSM results using the Göttingen Kp (solid curves) and the synthetic Kp* (dashed curves) for cases when the synthetic Kp* differs from Kp via standard deviations of 0.25, 0.50, 0.75, 1.0 and 1.25.

4. Dst INDEX

The Dst index, which is a measure of the ring current intensity, is an important input for both the MSM and the Magnetospheric Specification and Forecast Model (MSFM). The latter model requires the Dst index in real time, and we were asked to determine the status of this index. After discussions with the scientists at NOAA, the USGS, and the Air Force, we concluded that nobody was preparing an algorithm for calculating Dst in real time. We also concluded that Dst could be calculated in real time to the accuracy needed by the forecast models and recommended a procedure for doing this. In addition, we indicated what the quality of the magnetometer data must be in order to obtain the desired Dst accuracy. We were then tasked to develop an algorithm that would calculate real-time Dst from three magnetometer stations. The algorithm was developed and then delivered to the Air Force on 20 April 1998. A summary of our work on Dst is provided in what follows and a description of the algorithm that calculates the real-time Dst is given in the Appendix.

4.1 *Delivery and Implementation at RADEX Corporation*

DST_CALC Version 0.8 has been transferred to Radex Corporation as a fully operational program for producing a real-time Dst approximation using three magnetometer stations. Along with the working code a document detailing the DST_CALC program has been written and presented to Radex Corporation in hardcopy format. Vince Eccles of Space Environment Corporation visited Radex Corporation at Hanscom AFB (April 7-9) to test the implementation of DST_CALC to ensure its operation under the DST.COM program developed by Radex Corporation. DST_CALC operated as expected given the quality of the data collected by Radex from SSSG.

No changes to Version 0.8 DST_CALC were needed, so this version became the Version 1.0 DST_CALC and placed on CD-ROM for AFRL delivery. The DST_CALC document was updated to detail several important Version 0.8 items, and this document became the DST_CALC Version 1.0 documentation. The document was included on the CD-ROM and printed for delivery. The hardcopy document and CD-ROM were mailed to AFRL as a project deliverable.

4.2 *Specifics of Implementation and Testing in Dst.Com Environment*

DST_CALC Version 0.8 was placed on the Radex AlphaWorkstation which approximates the computer and database environment at the Space Forecast Center. DST_CALC was compiled and run on the computer. Compiler flag adjustments were required to make DST_CALC work properly within the DST.COM environment, but no other changes were required. Data into and Dst out from DST_CALC were examined to verify that DST_CALC is operating as expected. DST_CALC operated as it should --- given the data provided. No further programming is necessary for DST_CALC operation if data input is 'clean' data.

Even though there were no necessary changes to DST_CALC, a robust de-spiker was added to the DST_CALC program to remove spikes that appeared in the minute magnetometer data obtained by Radex Corporation from SSSG. Additionally, DST_CALC internal documentation was improved to provide aid in future changes to the coding when magnetometer stations are changed, added, or removed.

4.3 *Assessing the Dst Approximation using Real-Time Data*

For the entire development period for DST_CALC, the only data available for testing was clean magnetometer data for Honolulu, San Juan, and Guam. 'Realtime' data of the magnetometer stations were available at Radex Corporation for testing the DST_CALC code during the recent trip. The data were from the magnetometer stations at Guam, Honolulu, and San Juan. However, there were severe offsets within the data. The offsets occur at the date where the source of the data changed. The offsets in the data between the date of each source was surprising, but these could be fixed by applying a one-time offset value to the oldest data to match the newest data. However, within the data of each source (except the clean USGS data) there were data problems which severely compromise the Dst approximation produced by DST_CALC De-spiking the data (which DST_CALC V0.8 does by itself) was not sufficient to insure quality Dst approximations. DST_CALC was operating correctly when the USGS cleaned data was used. Data is assumed to be 'clean' with the exception of spikes which are removed by HMUS of SFC and/or DST_CALC.

4.4 *Data Problems*

The real-time data contains the following problems:

(1) Spikes (to be removed by HMUS and/or DST_CALC). Some of the flag files produced by HMUS were not present so the spike removal was incomplete. DST_CALC removed the remaining spikes.
(2) Permanent baseline offsets (not removed presently)
(3) Short-term baseline offsets (not removed presently)
(4) Short periods of large data drifts (not removed presently)

The real-time data problems 2-3 make DST_CALC a worthless Dst approximator. The problems can be addressed at one of three places --

(1) At the magnetometer stations through unknown quality assurance procedures,
(2) Within the HMUS program,
(3) Within the DST_CALC.

The first alternative is a worthy effort, but assuming that data will always clean of baseline offsets at the SFC computer is a dangerous assumption. The HMUS or DST_CALC program should be able to remove spikes and step-offsets from the data to insure that bad data will not damage the DST approximation.

The second alternative would be a suitable answer to the problem. HMUS already creates flags to identify spikes in the data. Additional software could be added to HMUS to identify offsets observed in the incoming minutely magnetometer data. The value of the offset must be communicated to the DST_CALC program through DST.COM. Cleaning the data via HMUS would also provide clean data for all programs using the magnetometer data. However, HMUS can only find offsets that occur within the 60 values or 12 values of the incoming magnetometer data, unless it loads a historical data set each time it runs. This may present a difficulty for HMUS in finding offsets that occur slowly or in correcting inaccuracies determinations in the observed offset.

The third alternative would be a suitable answer for the DST approximation alone. The benefits of placing data-cleaning software within DST_CALC are that offsets can be determined in the context of the historical magnetometer data 'remembered by DST_CALC. The offsets can be detected in minutely data using one to two days of minutely data. The determined offset found in the minutely data will have some remaining error in the data (what is a real variation and what is a long term offset?). The remaining error in the offset

correction can be corrected in the context of removing offsets in the hourly and daily averages. DST_CALC can be self-correcting in this manner to ensure the Dst approximation remains close to the real Dst. The negative of placing the data cleaning algorithms in DST_CALC is that the offsets are not communicated to other programs that use magnetometer data.

4.5 *Future Requirements for Quality Dst Approximation*

To detect offsets and drifts in the data, a thorough data characterization must be performed. Once the good minutely data and the data problems are characterized, then cleaning algorithm development can begin. There are at least several items which must be addressed in the future.

(1) Characterization of the good real-time data.
(2) Characterize observed problems.
(3) Develop filters for the removal of data problems. The filters must be applied to the incoming minutely H values as well as the long term data base. Some problems are readily apparent, others require a longer baseline for detection.
(4) Examine Dst Algorithm improvements for minimizing data-problem impacts.
(5) Improved RMS error to reflect data-problem impacts and real-time data characterization.
(6) Program and test data-filters and Dst algorithm improvements.
(7) Implement and test changes to programs DST_CALC and/or DST.COM and/or HMUS.

4.6 *Conclusion*

The DST_CALC program is working in the DST.COM environment and in ready for transfer to SSSG/SFC. The HMUS removes spikes, though this process may be imperfect. DST_CALC does remove all single point spikes. However, short term or permanent offsets to the data are not removed anywhere. The presence of an offset will compromise the DST approximation algorithms. The broadcast of the real-time Dst should be delayed until data cleaning procedures are implemented.

5. CONVECTION AND PRECIPITATION INPUTS

The plasma convection and particle precipitation patterns in the high-latitude ionosphere are important inputs for the three main forecasts models (IFM, TFM, MSFM). When these forecast models are run at the SFC, the convection and precipitation patterns will be needed in real time. However, the currently available patterns were statistical and it was not clear if these patterns would be adequate for Air Force needs. We were therefore tasked to determine the status of plasma convection and particle precipitation inputs. Our approach was to first determine how well the statistical models agree with satellite measurements on an orbit-by-orbit basis. If the agreement was not adequate, we would then recommend improvements to the statistical models. As a first step in our study, we had to acquire DMSP J4 particle data from the NDGC in Boulder, Colorado. We also acquired DMSP drift meter data from Peter Sultan at the Air Force Research Laboratory in Massachusetts.

In examining the statistical convection and precipitation models, we asked the following questions. What input quality is necessary for reliable IFM, TFM, and MSFM forecasts? What are the strengths and weaknesses of the statistical models in magnetically quiet, moderate, and active periods? If input improvements are proposed, will the improvements result in more reliable forecasts? If the statistical models fail some percentage of the time, is the failure with the models or with the indices that drive the statistical models (i.e., Kp and polar cap potential)?

The satellite data used in this study involved J4 particle data from four DMSP satellites (F8, F9, F10, F11). There were 730 Megabytes of binary data, which contained 5500 passes through the auroral oval. The dataset contained 15 quiet days, 8 disturbed days, and 3 storm onsets. Drift meter data from the SSIES instrument package were also available for these time periods.

5.1 *Precipitation Pattern*

We first compared the general statistical characteristics of our satellite measurements with those associated with the Hardy precipitation model. The energy flux and the mean position of the oval were in excellent agreement, as expected. Next, we compared the energy flux obtained from the Hardy model with the satellite measurements on an orbit-by-orbit basis. In some cases, we found that the shape and magnitude of the precipitation obtained from the Hardy statistical model were in remarkable agreement with the measurements (Figure 3), while in other cases there were substantial differences (Figure 4). Our study indicated that the broad statistical curve frequently misses the falloff of precipitation both poleward and equatorward of the precipitation peak. This can be true even when the statistical model correctly defines the peak magnitude. When the statistical model misses, the auroral oval frequently has sharp edges on either or both sides. For about 50% of the satellite crossings, the Hardy statistical model was in good agreement with the measurements. For about 40% of the crossings, the measured oval was narrower than the statistical oval. Finally, for about 10% of the crossings, the measured oval had an irregular appearance that was not consistent with the Hardy statistical oval.

The next issue we addressed was whether or not the differences between the Hardy statistical oval and the measured oval were important with regard to the accuracy of forecast model predictions. The Ionospheric Forecast Model (IFM) was used to study this issue. The model was run for a 3-day period (January 25-27, 1992) when four DMSP satellites were operating simultaneously. These data provided good coverage of the auroral oval and, therefore, the IFM could be run with a 'measured' oval in order to determine the

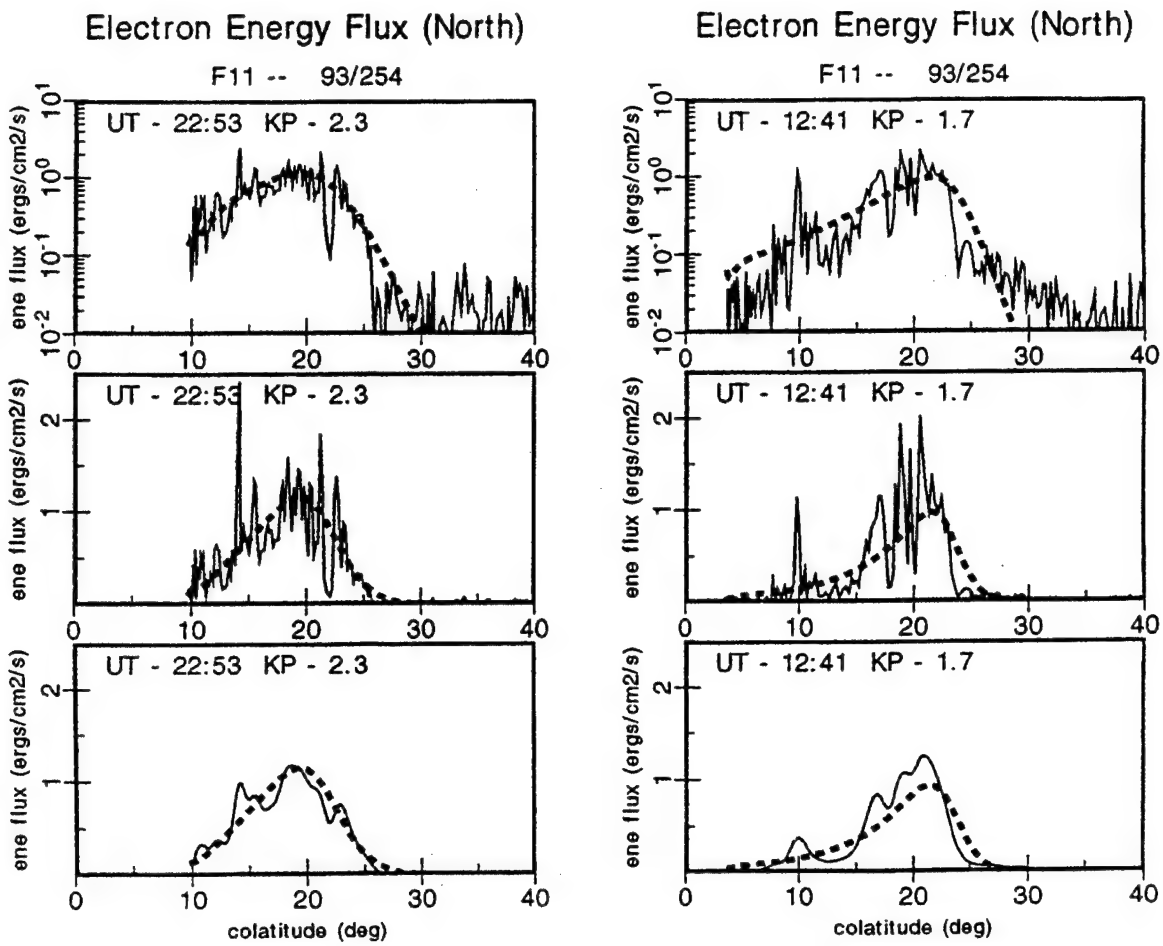


Figure 3. Comparison of the precipitating electron energy flux obtained from the Hardy statistical model with DMSP satellite measurements for cases when the agreement is excellent.

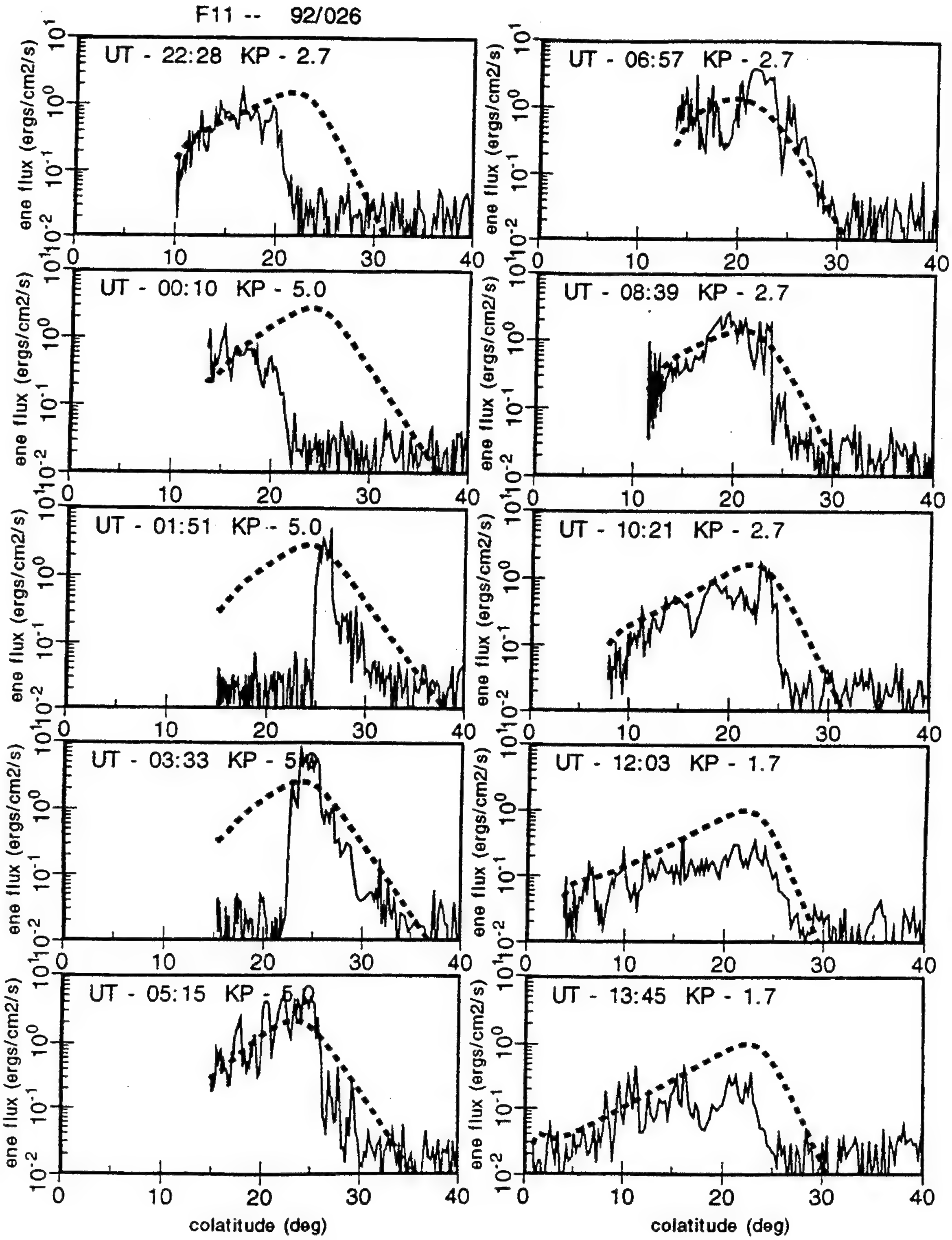


Figure 4. Same as Figure 3, but for cases when there are substantial differences between the Hardy statistical model and measurements.

ionosphere's sensitivity to 'real' precipitation. The IFM was also run with the Hardy statistical oval and the two model runs were then compared. We used a grayscale format to show the differences in the two IFM runs. A 50% difference was considered to be significant and, hence, a 3-level grayscale system was established:

$$f = \frac{Q_1 - Q_2}{Q_1} \qquad \begin{bmatrix} f < -0.5 & \text{white} \\ -0.5 < f < +0.5 & \text{gray} \\ f > +0.5 & \text{black} \end{bmatrix}$$

Figure 5 shows a representative comparison of the precipitation pattern obtained from the Hardy statistical model and the measured precipitation pattern. Note that the measured pattern was obtained by adjusting the Hardy model until it agreed with the simultaneous measurements from the four DMSP satellites. The measured patterns are called Hardy-J4. It is apparent from Figure 5 that there can be significant differences between the statistical and measured precipitation patterns. The differences in the precipitation patterns, in turn, lead to significant differences in the IFM electron densities at both the E-region (Figure 6) and F-region (Figure 7) altitudes. As a result of this comparison, we developed an algorithm that ingests DMSP J4 particle data and automatically modifies the Hardy statistical precipitation model so that it conforms to the real-time measurements. The algorithm was then delivered to the Air Force.

5.2 *Convection Pattern*

Our analysis of the plasma convection pattern was similar to that described above for the particle precipitation pattern. In particular, the drifts obtained from the Heppner-Maynard statistical model were compared, on an orbit-by-orbit basis, with those measured by the DMSP satellites. The F8 and F11 satellites were basically in dawn-dusk orbits, while the F9 and F10 orbital planes were along the 09 MLT and 21 MLT meridians. The horizontal component of the drift meter velocity that was perpendicular to the satellite track was used in our study. The equivalent velocity component was obtained from the Heppner-Maynard statistical model for comparison. Figure 8 shows typical comparisons of the statistical model results (dashed curves) and the measurements (solid curves) for quiet conditions, while Figure 9 shows representative comparisons for active magnetic conditions. In general, we found that even during quiet conditions there were systematic departures of the Heppner-Maynard statistical model from the measurements. For active magnetic conditions, large differences between the Heppner-Maynard model and the measurements were found. At times, there were significant differences in the locations of convection boundaries, and the flow reversal gradients that were measured were typically much greater than those obtained from the statistical model. These differences would be important with regard to how they would affect the IFM and TFM predictions. We therefore recommended that a better convection model should be produced.

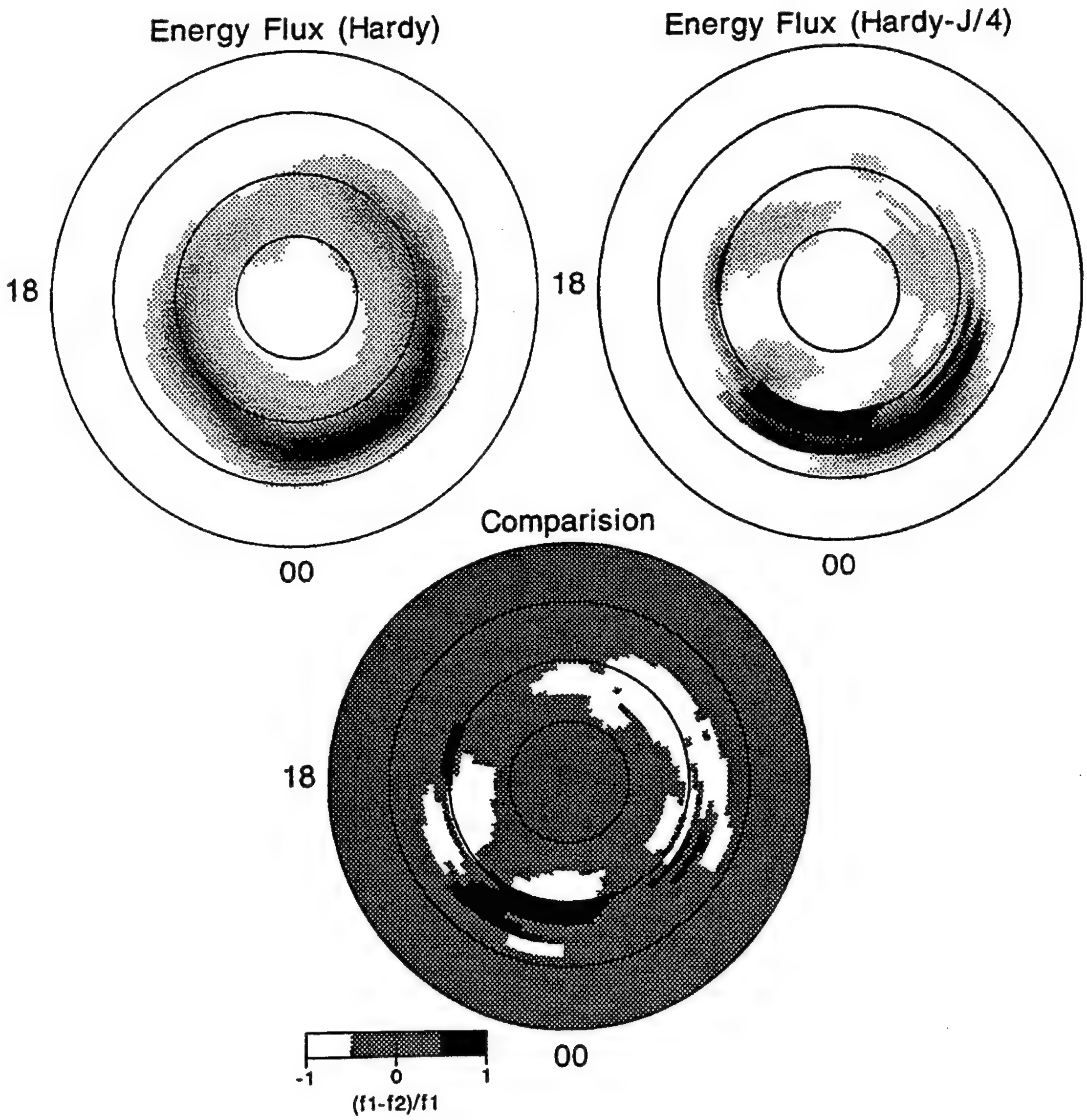


Figure 5. Precipitating electron energy flux obtained from the Hardy statistical model (upper-left panel) and the measurements (upper-right panel). The lower panel shows the difference between the two patterns.

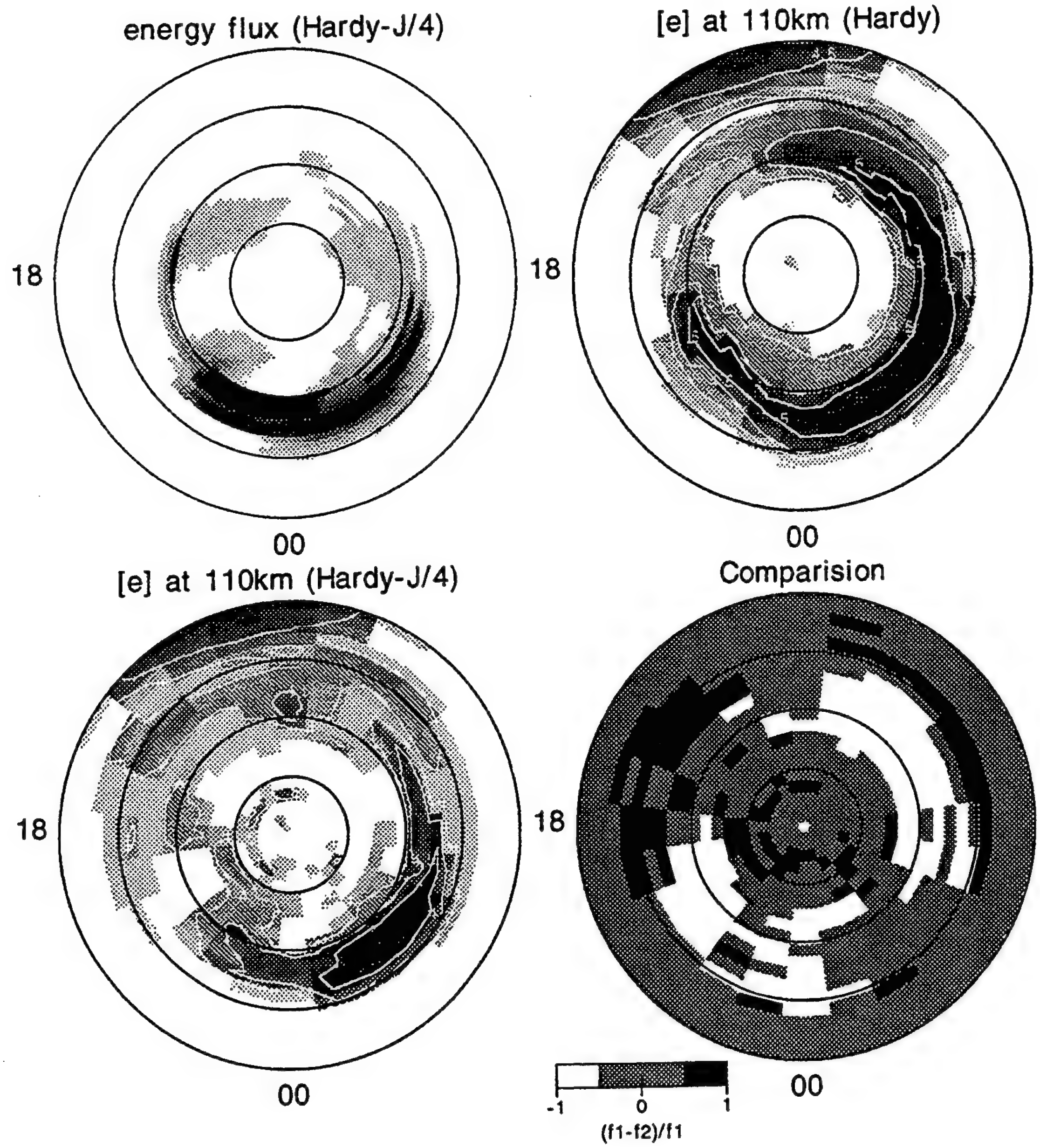


Figure 6. Electron densities at 110 km calculated with the Hardy statistical model (upper-right panel) and the measured precipitation (lower-left panel). The lower-right panel shows the differences in the electron densities calculated with the two precipitation patterns.

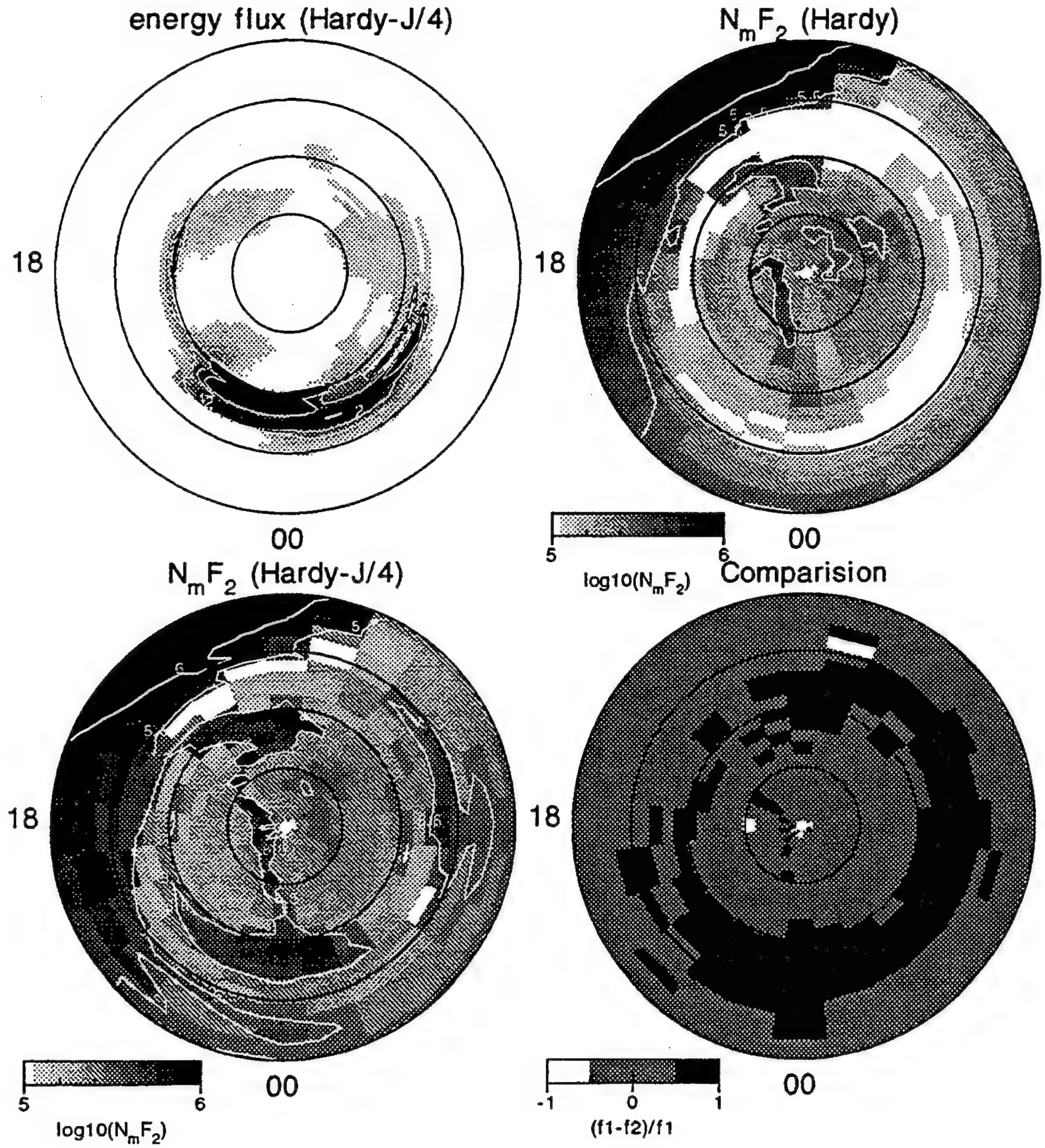


Figure 7. Same as Figure 6, except that the comparison is for N_mF_2.

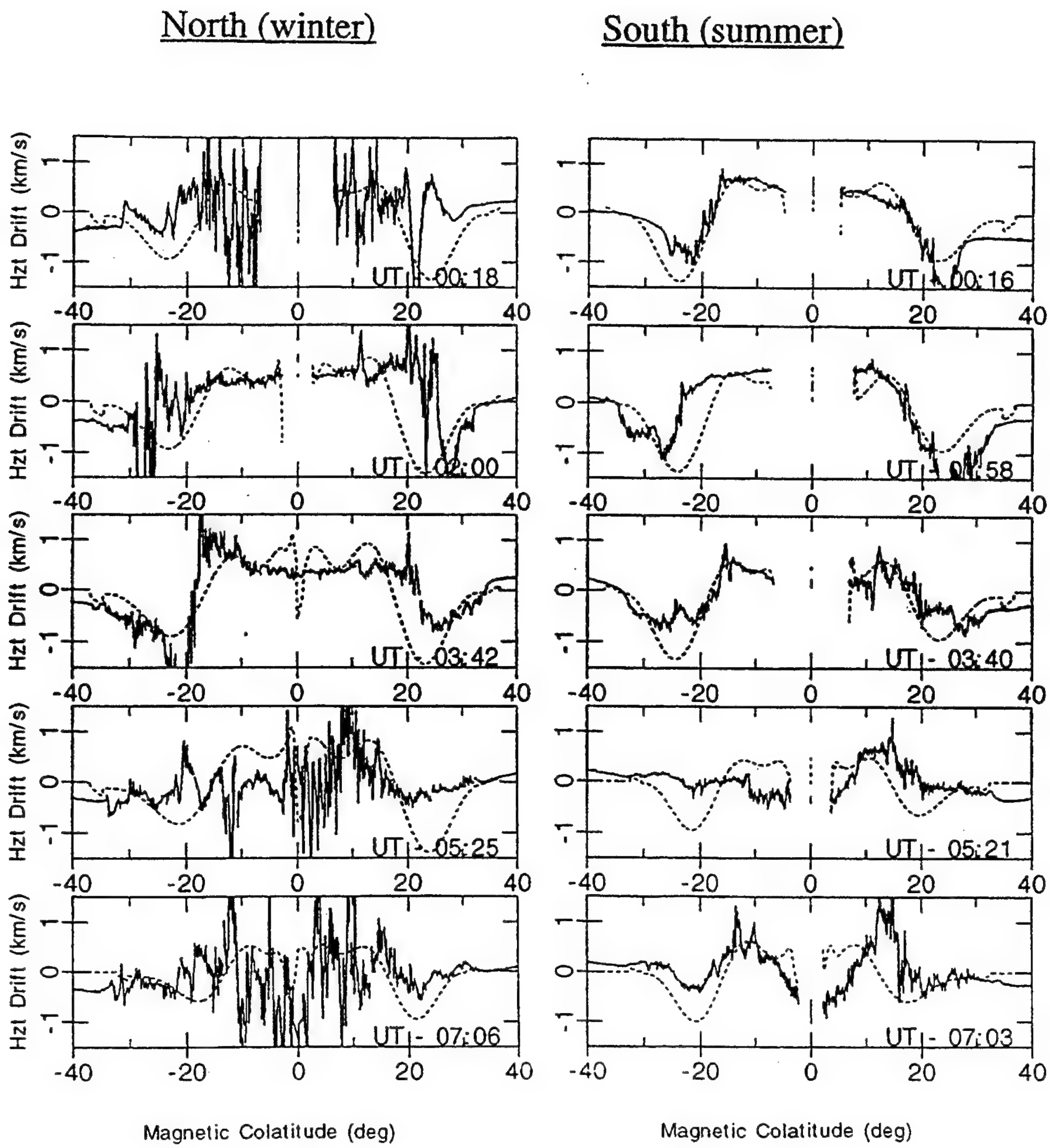


Figure 8. Comparison of horizontal plasma drifts obtained from the Heppner-Maynard convection model (dashed curves) with drifts measured by the DMSP satellites. The comparisons are for quiet magnetic conditions.

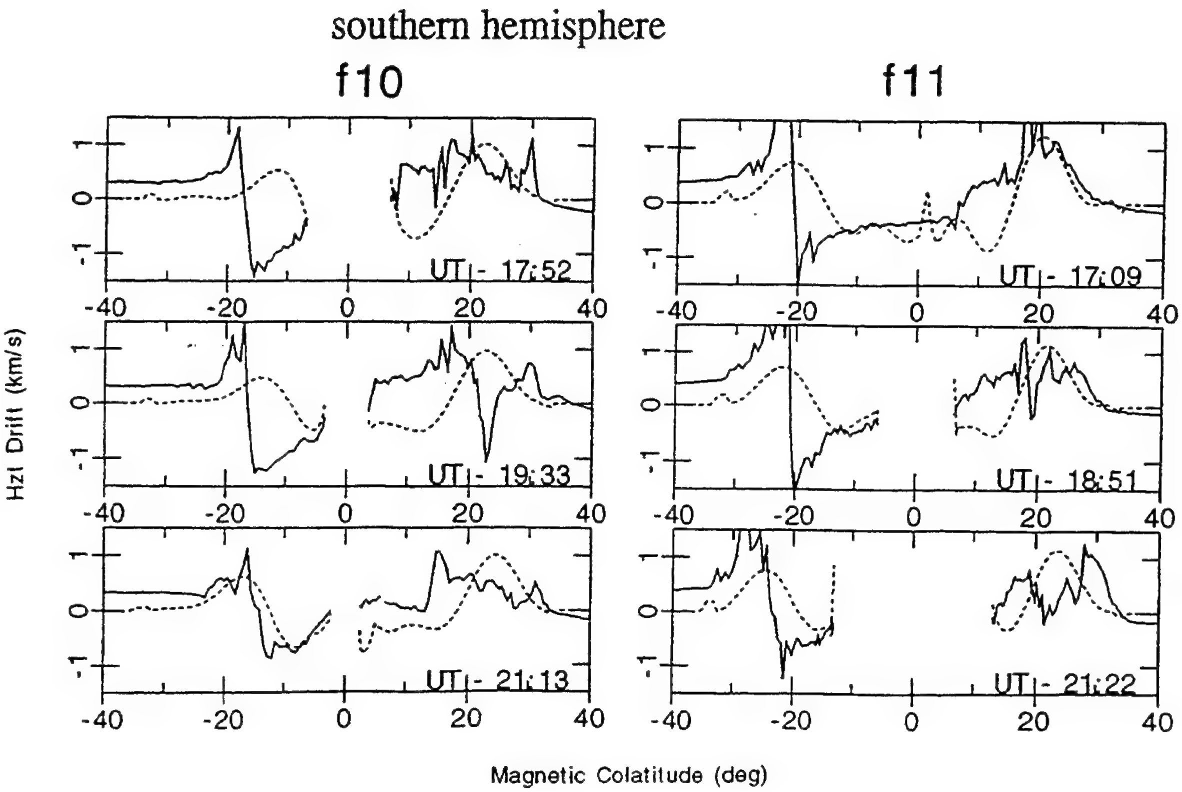


Figure 9. Same as Figure 8, except for active magnetic conditions.

6. QUALITY MANAGEMENT AND VERIFICATION

There are numerous data streams that arrive at the Space Forecast Center in near real-time, and several of these data streams are needed as inputs to the specification and forecast models. However, if the data quality is poor or if there are unexpected data gaps, the use of these data in the specification and forecast models could result in erroneous predictions. Therefore, an important aspect of an ISEM executive system involves an analysis of data quality. Specifically, if the data are to be automatically ingested in the specification and forecast models in real time, algorithms must be developed that can 'quality control' the data being ingested. One of our tasks was to study data quality issues.

We made an initial assessment of the quality of the datasets that flow into the Space Forecast Center, including the data from magnetometers, DMSP satellite instruments, the Ionospheric Measuring System (IMS), and the Digital Ionospheric Sounding System (DISS). Our analysis is given in the subsections that follow.

6.1 *Magnetometer Data*

The magnetometer data arrive at NOAA in real-time and then are sent to the SFC. We found that there were problems with the data. There were data gaps, data spikes of unknown origin, magnetic field components arbitrarily reversed, incorrect signs for some of the magnetic field components, and the UT of the measurements was uncertain by 12 minutes. These problems were reported to the Air Force, and subsequently they were corrected.

6.2 *DMSP J/4 Particle Data*

Orbital passes are used to identify the latitude of the equatorial boundary of electron precipitation. This is then mapped to midnight. A set of quality flags is associated with this analysis, which indicate the following: (1) How steep the latitudinal boundary is, ie, how accurately a boundary latitude can be determined; (2) How choppy the boundary is, ie, are their multiple choices for the boundary; and (3) The local time of the observed boundary, so that the uncertainty in mapping the boundary to midnight can be established (ie. a noon pass boundary is not well related to the required midnight boundary).

The ion and electron spectral information along each orbit pass does not have an associated quality flag. However, it is known that on occasions problems may arise due to spacecraft charging (winter-solar minimum - polar/auroral) and enhanced background due to solar cosmic ray events at times of major storms. Therefore, we recommended that quality flags be developed for the ion and electron spectral information.

6.3 *DMSP IES Data*

Software to extract the scientific parameters from this instrument had been developed. However, the issue of generating a set of quality flags was still to be determined. We identified several quality factors and recommended that the following steps be taken:

- Checks are run for obviously bad data. When identified, these are replaced with null values. The SFC algorithms that use the data streams must be able to recognize the null quality flag.

- Data files can, on rare occasions, be "fill" files. These are not identified or

flagged, hence this needs to be done.

- When light ions (H^+ and/or He^+) are dominant, the planar ion trap, giving n_e and Δn_e, and the drift meter, giving two components of the ion drift, are both contaminated. However, the retarding potential analyzer can be used to measure ion composition, and hence, a flag can be established that indicates when such periods are encountered.

- The electron probe analysis depends strongly on spacecraft potential. Data concerning the spacecraft potential is available, hence a quality flag should be developed.

6.4 *DMSP E-Field Data*

A second level of processing takes the DMSP IES drift data and computes information about the global electric field pattern. The algorithm was developed by M. Hairston (UTD) in conjunction with F. Rich (PL). The key scientific parameter is the cross polar cap potential which is obtained by integrating the electric field along the DMSP orbit pass. Quality flags have been developed for this analysis. Our assessment is as follows.

- When an orbit is incomplete, ie, northern pass up to Thule (the ground station), at time of transmission the satellite may not have completed the polar pass, an "extrapolation" is needed. Flags indicate what was done in the analysis. <u>Note</u> that a partial orbit is important because it is real-time data.

- If the potential drop is less than 30 kV, the inference of the type of convection pattern is difficult. This condition can be flagged.

- The local times at which the orbit planes cut through the convection pattern are critical in deciding how well the potential drop and pattern have been inferred. This information can be used to establish the overall quality of the procedure.

- At this time of our assessment, the algorithm did not use any quality flag information from the IES data reduction. It assumed the original drift data are of good quality.

6.5 *Ionospheric Measuring System Data*

Although, at the time of our assessment, data were not sent to the SFC, the dual-frequency GPS TEC system was being field evaluated. The IMS receiver at Otis ANGB, Massachusetts was being interrogated remotely from PL by Greg Bishop. We found the following:

- The IMS AWN data packet has extensive data quality information, including; which satellites were used, how many samples form the average value, the strength of the signal at each frequency, and which tracking mode was used.

- Calibration of the GPS derived TEC is a crucial aspect of using these data. In addition to the IMS calibration, software has been developed (Greg Bishop, PL) that takes 24 hours of data from a GPS/TEC ground station and calibrates the data to $\pm$ 1 TEC unit. Hence, the quality of TEC data from different sources

can be established.

6.6 *DISS Data*

Data from the DISS stations are transferred over the AWN network to the SFC. These data packets are transferred in near real-time (<15 minutes). These data are the key real-world input for the PRISM ionospheric specification model. At the time of our assessment, the quality of the DISS data packets was being evaluated and improved in the following manner:

- Terry Bullett (PL) is responsible for the operational DISS stations and is actively concerned about quality.

- Ray Conkright (NGDC) has been tasked to establish how well the remote software is inverting the real time ionogram to scientific parameters in the AWN data packet. The software is based on the ARTIST ionogram inversion program. Based upon this study and the experience gained in real-time ionogram analyses, improvements in the algorithm are expected.

- The DISS-AWN data packet has a fixed structure and limited expansion capability. It contains very little quality information. This is an area for future upgrading.

- The DISS-AWN data packet contains information that allows for the reconstruction of the EDP as well as the ionogram used by ARTIST. These data can also be extracted from the specification and forecast data sets by an ISEM executive system. Hence, an executive system can carry out a quality control, a compatibility check of data and model.

6.7 *Quality Management and Verification Approach*

We studied the advantages and disadvantages of two different Quality Management and Verification (QMV) approaches, and our analysis is given in a separate report attached to this document. The two approaches we considered were: (1) Build the QMV software inside the product software; and (2) Develop separate QMV software to evaluate the quality of data products. We proposed the second approach to the Air Force, for the reasons given in Table 2. We then constructed and tested a prototype at our company. The prototype was based on simple indices, such as Kp, Ap, SSN, and F10.7. In what follows, only the results for Kp are discussed.

Table 2. Advantages of Independent QMV utilities.

1. Requires only a knowledge of the software product, not how the product was generated or its possible failure modes.

2. Consists of a sequential set of tests; hence a fixed number of operations, which means no unspecified CPU requirements.

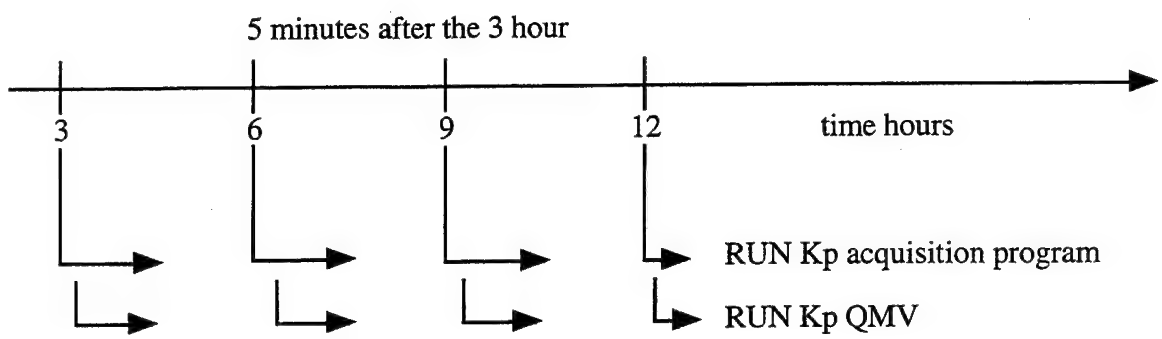
3. Output is a fixed record.

4. QMV report deals with how well the real-time utilities output met the design specification.

5. At this level QMV is not carrying out "damage control". It provides the necessary information to readily establish the degree of damage in reports as simple as on number/month if desired.

QMV Software

- The basic NOAA SEL Kp and other indices are updated at 3-hour intervals. SEC acquires these data and stores them as the current geophysical indices file.
- A program to check this real-time acquisition runs automatically on completion of the real-time data acquisition program.
- Execution timeline is 5-minutes after the 3-hour mark,



- Once daily and once monthly report generators are run a few minutes after the end of a day or month. These produce daily or monthly QMV reports.
- Examples of the March 1996 and first 21 days of April 1996 monthly reports are shown in the pages that follow. These files are 2 K bytes long, and a total of 12 monthly files exist, a total of 24 K bytes.

Monthly QMV Summary for March 1996

```
 03.1996 SUMMARY PAGE
**********************************************************
     SSN_00000000
     --------------------------------
     All OK.                        |  45 of  56.
     No file found - sgas.txt.      |  11 of  56.
     --------------------------------
     Success rate: 80%
**********************************************************
     F10_00000000
     --------------------------------
     All OK.                        |  45 of  56.
     No file found - sgas.txt.      |  11 of  56.
     --------------------------------
     Success rate: 80%
**********************************************************
     KP3_00000000
     --------------------------------
     All OK.                        | 142 of 222.
     Bad DOY/time.                  |  10 of 222.
     Ap Kp mismatch.                |   1 of 222.
     No file found - MAhr.txt.      |  69 of 222.
     --------------------------------
     Success rate: 63%
**********************************************************
     AP3_00000000
     --------------------------------
     All OK.                        | 142 of 212.
     Ap Kp mismatch.                |   1 of 212.
     No file found - MAhr.txt.      |  69 of 212.
     --------------------------------
     Success rate: 66%
**********************************************************
     Week00000000
     --------------------------------
     No files found.                |   5 of   5.
     --------------------------------
     Success rate:  0%
**********************************************************
     KP__00000000
     --------------------------------
     No file found - mada.txt.      |  11 of  11.
     --------------------------------
     Success rate:  0%
**********************************************************
     AP__00000000
     --------------------------------
     No file found - mada.txt.      |  11 of  11.
     --------------------------------
     Success rate:  0%
**********************************************************
```

Monthly QMV Summary for April 1996

```
 04.1996 SUMMARY PAGE
*************************************************************
      KP3_00000000
      --------------------------------
      All OK.                         | 112 of 252.
      Bad DOY/time.                   |  20 of 252.
      Bad jump in Kp.                 |   1 of 252.
      No file found - MAhr.txt.       | 119 of 252.
      --------------------------------
      Success rate: 44%
*************************************************************
      AP3_00000000
      --------------------------------
      All OK.                         | 112 of 232.
      Bad jump in Ap.                 |   1 of 232.
      No file found - MAhr.txt.       | 119 of 232.
      --------------------------------
      Success rate: 48%
*************************************************************
      SSN_00000000
      --------------------------------
      All OK.                         |  44 of  60.
      No file found - sgas.txt.       |  16 of  60.
      --------------------------------
      Success rate: 73%
*************************************************************
      F10_00000000
      --------------------------------
      All OK.                         |  44 of  60.
      No file found - sgas.txt.       |  16 of  60.
      --------------------------------
      Success rate: 73%
*************************************************************
      KP__00000000
      --------------------------------
      No file found - mada.txt.       |  16 of  16.
      --------------------------------
      Success rate:  0%
*************************************************************
      AP__00000000
      --------------------------------
      No file found - mada.txt.       |  16 of  16.
      --------------------------------
      Success rate:  0%
*************************************************************
      Week00000000
      --------------------------------
      No files found.                 |   5 of   5.
      --------------------------------
      Success rate:  0%
*************************************************************
```

Analysis of Kp Monthly Reports

- March 1996 had three Kp messages and their frequency of occurrence, while April 1996 had five Kp messages.
- ALL OK, an obvious message occurred 63% of the time in March and 44% in April.
- NO FILE FOUND, also an obvious message which occurred 31% of the time in March and 47% of the time in April.
- BAD DOY/TIME, the date and/or time has not gone forward since the last file was read. One expects three hour increments. Out of 133 successful Kp reads from NOAA, 20 cases of wrong DOY/time were found (4%) in March and (8%) in April.
- Ap-Kp MISMATCH, one such event occurred in March. This event indicates that the statistical log (Ap) to Kp relationship was not satisfied. (Note the relationship and standard deviations were based on 40 years of Ap-Kp data).
- BAD JUMP IN Kp, one such event occurred in April 1996. This represents the case where a Kp value changed by more than +/- 4.

Response to Summary

- The SEC networking system is rather cheap! It appears to work about 50 to 60% of the time.

  Hence SEC is "forecasting" a lot of 3-hour real-time indices.

- BAD DOY/TIME. Search, the daily reports for more information on these errors. Note daily reports are discussed next.

  Unix is ideal for this, just do a grep for the error message on the daily QMV messages and list these lines. ? is this an SEC bug, probably yes.

- BAD JUMP IN Kp. Searched the daily files and found this to have occurred at 21:09 MST on April 19, 1996. A look at the daily file showed the jump to be from 6^- to 1^+. Looking at the activity around that period, it is probably OK.
- Ap-Kp MISMATCH. This occurred in March. Unfortunately, the daily record files for March 1 through 21 had been overwritten with the newer April 1 to 21 data. SEC "operators"! don't check summaries very often.
- The BAD DOY/TIME, Ap-Kp MISMATCH errors, and warnings should have alerted the operator.

Grep on Bad DOY/Times

- The date and time of the occurrences are reported.
- The proceeding times are all the 7th Kp in the day.
- The subsequent date and time are definitely wrong.

```
QVM_daily02.stat: KP3_00100000 - 02.04.1996 00:09:48 | Bad DOY/time.          91.88  1.92
QVM_daily06.stat: KP3_00100000 - 06.04.1996 00:09:12 | Bad DOY/time.          95.88  5.92
QVM_daily10.stat: KP3_00100000 - 09.04.1996 00:15:09 | Bad DOY/time.          98.88  1.92
QVM_daily10.stat: KP3_00100000 - 09.04.1996 03:13:28 | Bad DOY/time.          98.88  1.92
QVM_daily10.stat: KP3_00100000 - 09.04.1996 06:12:52 | Bad DOY/time.          98.88  1.92
QVM_daily10.stat: KP3_00100000 - 09.04.1996 09:13:13 | Bad DOY/time.          98.88  1.92
QVM_daily11.stat: KP3_00100000 - 11.04.1996 00:06:36 | Bad DOY/time.         100.88  3.92
QVM_daily11.stat: KP3_00100000 - 11.04.1996 03:09:14 | Bad DOY/time.         100.88  3.92
QVM_daily14.stat: KP3_00100000 - 14.04.1996 00:09:13 | Bad DOY/time.         103.88  6.92
QVM_daily18.stat: KP3_00100000 - 18.04.1996 00:09:53 | Bad DOY/time.         107.88  3.92
QVM_daily20.stat: KP3_00100000 - 20.04.1996 00:10:24 | Bad DOY/time.         109.88  5.92
QVM_daily23.stat: KP3_00100000 - 23.04.1996 00:06:35 | Bad DOY/time.         112.88  1.92
QVM_daily23.stat: KP3_00100000 - 23.04.1996 03:09:26 | Bad DOY/time.         112.88  1.92
QVM_daily24.stat: KP3_00100000 - 24.03.1996 00:10:44 | Bad DOY/time.          82.88  6.92
QVM_daily26.stat: KP3_00100000 - 26.03.1996 00:09:12 | Bad DOY/time.          84.88  1.92
QVM_daily27.stat: KP3_00100000 - 27.03.1996 00:09:13 | Bad DOY/time.          85.88  2.92
QVM_daily30.stat: KP3_00100000 - 30.03.1996 00:06:51 | Bad DOY/time.          88.88  5.92
QVM_daily30.stat: KP3_00100000 - 30.03.1996 03:07:19 | Bad DOY/time.          88.88  5.92
QVM_daily30.stat: KP3_00100000 - 30.03.1996 06:07:19 | Bad DOY/time.          88.88  5.92
QVM_daily30.stat: KP3_00100000 - 30.03.1996 09:09:13 | Bad DOY/time.          88.88  5.92
QVM_monthly03.stat:    Bad DOY/time.          | 10 of 222.
QVM_monthly04.stat:    Bad DOY/time.          | 20 of 252.
```

19 April 1996 Daily Summary

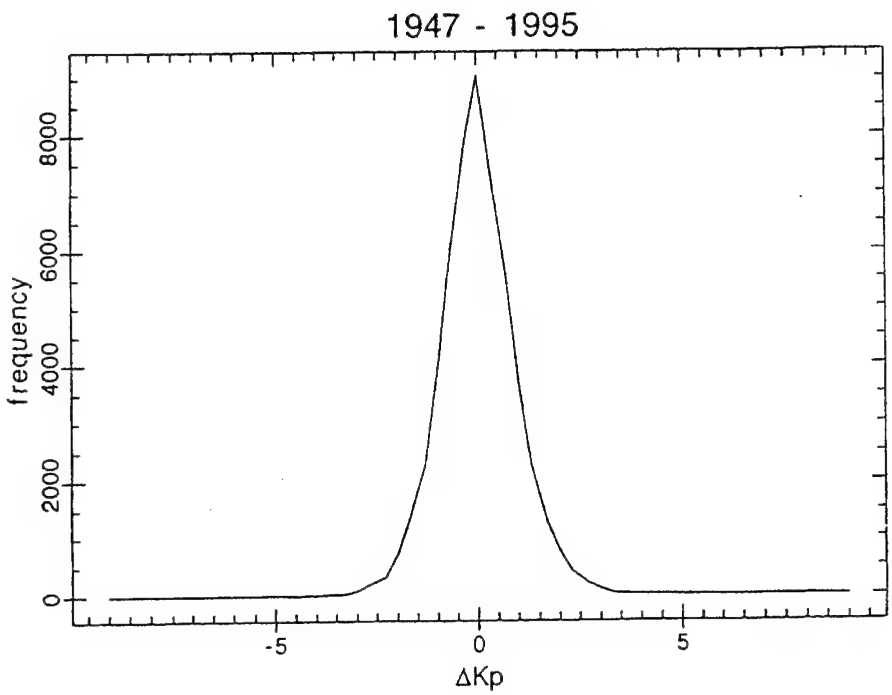
- Report showing the Bad jump in Kp.
- Prior value was 6-, while the new value was 1+. The absolute difference is 5.
- QVM_daily19.stat file:

```
KP3_00000000 - 19.04.1996 00:09:57 | All OK.                     3.7
AP3_00000000 - 19.04.1996 00:09:57 | All OK.                    25.0
SSN_00000000 - 19.04.1996 01:10:06 | All OK.                    14.0
F10_00000000 - 19.04.1996 01:10:06 | All OK.                    69.0
KP3_00000000 - 19.04.1996 03:09:23 | All OK.                     3.7
AP3_00000000 - 19.04.1996 03:09:23 | All OK.                    20.0
SSN_00000000 - 19.04.1996 05:38:00 | All OK.                    29.0
F10_00000000 - 19.04.1996 05:38:00 | All OK.                    70.0
KP3_00000000 - 19.04.1996 06:09:18 | All OK.                     4.3
AP3_00000000 - 19.04.1996 06:09:18 | All OK.                    31.0
KP3_00000001 - 19.04.1996 09:14:48 | No file found - MAhr.txt.
AP3_00000001 - 19.04.1996 09:14:48 | No file found - MAhr.txt.
KP3_00000001 - 19.04.1996 15:13:29 | No file found - MAhr.txt.
AP3_00000001 - 19.04.1996 15:13:29 | No file found - MAhr.txt.
KP3_00000000 - 19.04.1996 18:09:22 | All OK.                     4.0
AP3_00000000 - 19.04.1996 18:09:22 | All OK.                    28.0
KP3_00001000 - 19.04.1996 21:09:24 | Bad jump in Kp.             1.3   5.
AP3_00001000 - 19.04.1996 21:09:24 | Bad jump in Ap.             5.0  65.0
```

- An eight-character error flag identifies, by a 0 being replaced by a 1, the specific error.

Change in Kp check

- Examining ΔKp for 1947-1995 provides flag for Bad Kp.



ΔKp	frequency	ΔKp	frequency
-6.000000	0.0000000E+00	0.0000000E+00	9025.000
-5.700000	0.0000000E+00	0.3000000	7503.000
-5.300000	0.0000000E+00	0.7000000	5591.000
-5.000000	1.000000	1.000000	3724.000
-4.700000	0.0000000E+00	1.300000	2256.000
-4.300000	6.000000	1.700000	1260.000
-4.000000	5.000000	2.000000	773.0000
-3.700000	18.00000	2.300000	406.0000
-3.300000	28.00000	2.700000	199.0000
-3.000000	76.00000	3.000000	105.0000
-2.700000	183.0000	3.300000	33.00000
-2.300000	307.0000	3.700000	16.00000
-2.000000	700.0000	4.000000	13.00000
-1.700000	1339.000	4.300000	8.000000
-1.300000	2293.000	4.700000	6.000000
-1.000000	3913.000	5.000000	1.000000
-0.7000000	5757.000	5.300000	1.000000
-0.3000000	7902.000	5.700000	0.0000000E+00
		6.000000	0.0000000E+00

Error Codes

- Error Word has 8 characters, when 0 it implies that this error has not occurred, when 1 implies this specific error occurred.

0 0 0 0 0 0 0 0

Data file not found
File found but stale
Bad Value in data
Bad jump in Value
Data-to-Data Mismatch
Doy/Time out of sequence
Future use
Future use

- Error Word specifies data.

KP3_	3 hour Kp
AP3_	3 hour Ap
AP_ _	Daily Ap
KPS_	ΣKp
SSN_	Sunspot #
F10_	F10.7
F10A	F10.7a

Conclusions and Summary of this QMV Prototype

- Total prototype QMV software is 2000 lines. QMV report file sizes:

12 monthly files	= 12 x 2 K bytes	=	24 K bytes
31 daily files	= 31 x 1.5 K bytes	=	47 K bytes
total file space	= 70 K bytes		

- CPU on DEC ALPHA is less than a second/usage. In 24 hours this total is less than 30 seconds.

- We very quickly learned that a Mountain Standard Time day is not the same as a Universal Time day.

- This leads to problems if software is used in other time zones, and in computing daily averages. *We recommend all work be done in UT.*

- This software has been running at our company since 4 March 1996, no operator time is needed, one can ignore the reports. They do not pile up and collect dust, they get overwritten. Daily and/or monthly printing of short reports can be done. (Presently, 377 pages/year)

- Weekly reporting was not done. We could not readily define what a week is, i.e., at start of the month and end, how do you choose the week boundaries? Do you mix the end of one month with the start of next month?

APPENDIX

DESCRIPTION OF DST_CALC:
REALTIME DST APPROXIMATION PROGRAM

by Vince Eccles

April 1998

Space Environment Corporation
399 North Main, Suite 325
Logan, UT 84321
(435) 752-6567

TABLE OF CONTENTS

1. INTRODUCTION

2. DST_CALC USAGE

3. GENERAL PHILOSOPHY OF DST ALGORITHMS

3.1. Approximation of Dst from a Single Mag. Station.
3.2. Secular Variation.
3.3. Quiet Day Variation.
3.4. Dst from ΔH through Regression Coefficients.
3.5. Single RDst Approximation.

4. PROGRAM STRUCTURE
4.1. Introduction.
4.2. Input Modules.
4.2.1. READ_IN_FILES Module.
4.2.2. UPDATE_DBS Module.
4.2.3. READ_TRENDS Module.
4.3. Calculation Modules.
4.3.1. DO_TRENDS Module.
4.3.2. DO_DST Module.
4.3.3. DO_DST_ERROR Module.
4.4. Output Module.
4.4.1. WRITE_OUT_FILES Module.

5. PUBLIC DOMAIN SOFTWARE USED IN DST_CALC
5.1. NETLIB
5.1.1. FFTPACK
5.1.2. NAPACK
5.2. JVE UTILITIES
5.2.1. mynaJVE.f
5.2.2. dateJVE.f
5.2.3. charJVE.f

6. ADDING OR REMOVING OR CHANGING A STATION IN DST_CALC

1. INTRODUCTION

The Air Force Research Laboratory has funded this project to produce an approximation of the Dst index in realtime. The Space Weather Squadron (55th SWXS) has a defined need for a realtime Dst to drive models of the space environment. DST_CALC is the program that produces the Dst approximation (RDST) from the de-spiked magnetometer data available in near realtime at the Space Forecast Center. This document covers the usage of DST_CALC, the general philosophy of the algorithms, and the program structure. Other aspects of DST_CALC, i.e., interfacing of DST_CALC with Space Forecast Center databases or detailed description of the science behind the Dst approximation algorithms, are covered in other documents associated with the overall Dst project.

2. DST CALC USAGE

DST_CALC is a program written in FORTRAN 77 with industry standard extensions. It has been written to run under most operating systems including UNIX and VMS. It runs silently, that is, without screen prompts or messages, unless there is a program error. If a common error does occur, then a subroutine passes information to the error message handler on the Space Forecast Center computers.

DST_CALC requires the presence of input files and database files. If these files are not present an error message is sent and the program is terminated. Table 1 lists necessary the input files and the database files. The format of the input files are discussed in the Radex SVS document. The historical data files are defined in the following sections.

TABLE 1. Necessary files for DST_CALC.

Input files:

- DST_H_IN.DAT
- DST_KPF10_IN.DAT (only on first run of the UT day)
- DST_SUCCESS.MSG

Historical data files: xxxxxx is station number

- DST_CALC_TODAYxxxxxx.DAT
- DST_CALC_YESTERDAYxxxxxx.DAT
- DST_CALC_HOURLYxxxxxx.DAT
- DST_CALC_DAILYxxxxxx.DAT
- DST_CALC_KP.DAT

DST_CALC is called by DST.COM. It reads the necessary files, makes the Dst approximation for each minute of magnetometer data contained in the input files, then outputs the Dst and error estimate for each minute contained in the input files. DST_CALC outputs the DST approximation in DST_OUT.DAT. DST_CALC also output the solar quiet variation of the H component of the magnetic field at each station used in calculating DST. The output file is called DST_HSQ_OUT.DAT. The secular variation of H at each station is output in DST_HSV_OUT.DAT. All output files are ascii and have formats defined in the Radex SVS documentation of DST.COM.

DST_CALC assumes that there are no greater than 60 minutes of data and that the data input will not cross UT hour boundaries. DST_CALC also assumes that there is a particular set of magnetometer stations providing data. The first version of DST_CALC assumes three stations; San Juan, Honolulu, and Guam. If this set in reduced or increased in number or if a station is replace by another site's station then DST_CALC must be changed. The necessary alterations are limited to include files and one subroutine. These are documented in sub-section 4.5.

3. GENERAL PHILOSOPHY OF DST ALGORITHMS

3.1 Approximation of Dst from a Single Magnetometer Station

This section does not contain a comprehensive description of the algorithms. Instead, a general outline of the algorithms within DST_CALC is provided to aid in understanding the discussion of the program structure below.

The Dst index is an indicator of ring current strength around the earth. Solar wind modification of the earth's environment produces an increase in the ring current strength which decays over several days. The ring current generates an axial magnetic field that is observed by mid-latitude magnetometer stations as a small deviation superimposed on the much larger earth field measurement. The Dst index is based on the horizontal component of the magnetometer measurements, H. The H component from several geographically spaced stations are used to remove sector dependencies. The Dst algorithms presented herein are suitable for producing a Dst approximation from a single station measurement. Several single station Dst estimates are averaged to improve the Dst approximation and provide redundancy if a station is temporarily off-line.

There are several contributors to H. The main component of the earth's field strength is large and slowly varying when compared with magnetic storm variations. The earth's slowly varying field strength is called the Secular

Variation, SV. The horizontal field portion of the SV will be called HSV. There is also a daily variation of the H value arising from horizontal electric currents in the ionosphere. The ionospheric currents have daily repeatable structure because they are driven by solar energy deposition in the upper atmosphere. This portion of H is approximately reproducible from day-to-day if magnetic storm effects and substorm effects are removed. The day-to-day, quiet-condition, repeatable variation is called the Solar Quiet contribution, SV, or HSV for the horizontal component. Both the Secular Variation, HSV, and the Solar Quiet variation, HSQ, are different for stations with different geographic locations. The following description of the Dst algorithms focus on removing the HSV and then the HSQ from H to obtain the stations storm induced deviations, ΔH, that is,

$$\Delta H_s = (H_s - HSV_s - HSQ_s) \tag{1}$$

where s is the station indices.

Because Dst is a measure of the ring current's strength and the ring current produces an axial magnetic component, the Dst algorithm must account for the latitude of the station.

$$Dst_s' = \Delta H_s / \cos(\lambda_\sigma) \tag{2}$$

The Dst_s' has a prime to indicate that a further refinement is possible beyond Eq (2). Each station responds to the ring current disturbance according to its Local Time (or UT and Geolongitude of the measurement). Thus, a linear relation has been used to map a single station measurement of Dst_s' into a truer Dst approximation. This approximate, realtime Dst will be referred to as RDst.

$$Dst_s \cong RDst_s = A_s(UT)*Dst_s' + B_s(UT) \tag{3}$$

where A_s and B_s are regression coefficients based on historical data of a single station and the real Dst. The algorithm was provided by Robert McPherron of Space Environment Corporation and is detailed in a report provided by Geoff McHarg of the Air Force Academy.

3.2. Secular Variation Calculation.

The secular variation of the magnetic field's horizontal component, HSV_s, is the slow changing component of H_s caused by the earth's intrinsic field. Each station sees a different secular variation. HSV_s is essentially constant for time periods

shorter than a month. HSV_s is found by determining a best-fit polynomial on 10 years of quiet day averages of H_s for each solar rotation, 27 day period. That is, first, divide 10 years of daily average H_s values into 27 day periods; second, averaged the top 20% of the daily average values in each rotation to remove storm days; third, determine the 5th order polynomial approximation to the 10 year trend in H_s (Figure 1). This 5th order polynomial will be the best fit HSV_s trend to be removed from H_s.

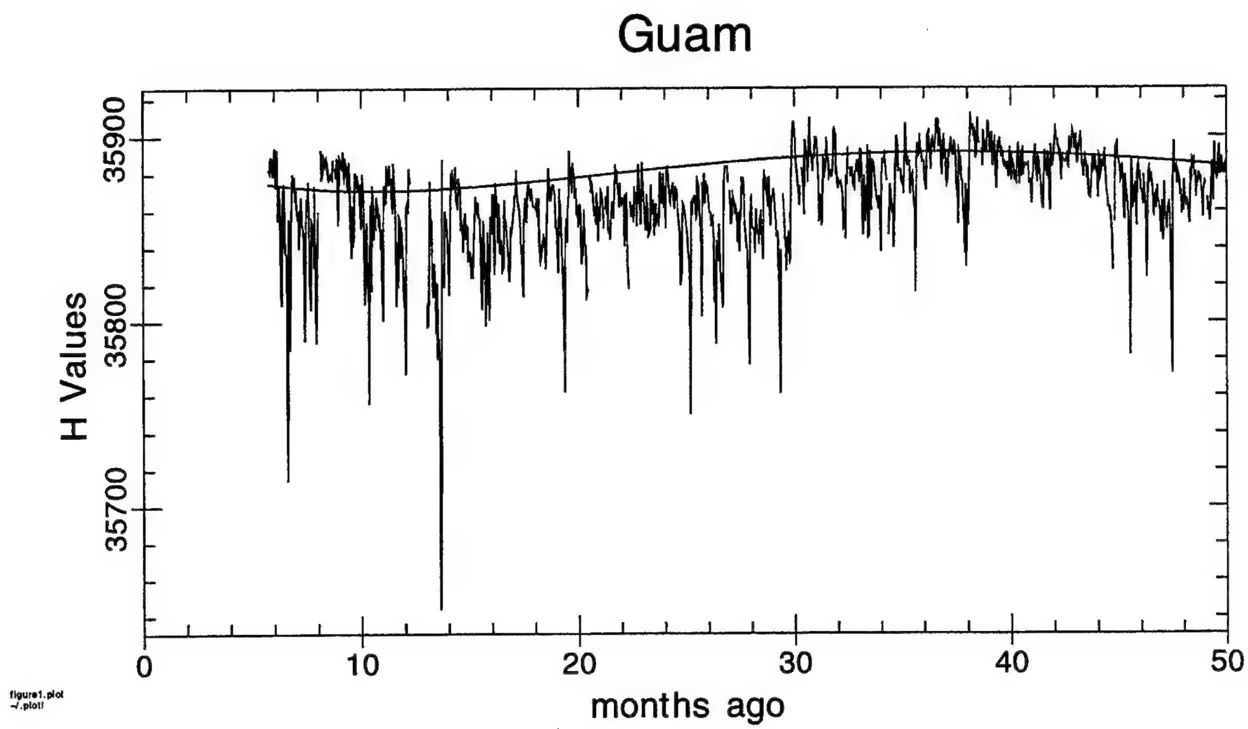


Figure 1. Secular variation of the H component from the Guam magnetometer station.

3.3. Solar Quiet Variation Calculation.

The solar quiet variation of the horizontal component, HSQ_s, is the daily variation caused by the neutral-wind-driven currents in the overhead ionosphere. It is assumed the sun alters conductivity and winds in a regular fashion over each station and an average daily trend can be extracted from the observations. To do this, hourly averages of the previous 365 days are stored to provide a seasonally periodic data set that accommodates FFT filtering. To obtain the quiet day one must first, remove the secular variation; second, remove storm influences; third, obtain the repeatably-smooth daily variation in H_s by FFT analysis. The secular

variation is removed from the year of historical H_s hourly averages by subtracting the best-fit 5th-order polynomial. Magnetic storm and recovery phase trends are removed by obtaining a cubic spline fit to the low points in the solar quiet current trend which occurs near local midnight for each station (Figure 2). This midnight cubic spline fit is subtracted from the H_s-HSV_s remnant. To insure the storm effects do not influence the quiet day, determinations times during high Kp, i.e., during storms and sub-storms, are replaced with BAD flags. All that is left in the year long historical hourly averages is a noisy quiet day component. There is still considerable day-to-day variation in the signal supposedly due to smaller non-storm ring current variations. FFT filtering is used to obtain a best quiet day variation for each station, HSQ_s (Figure 3 & 4).

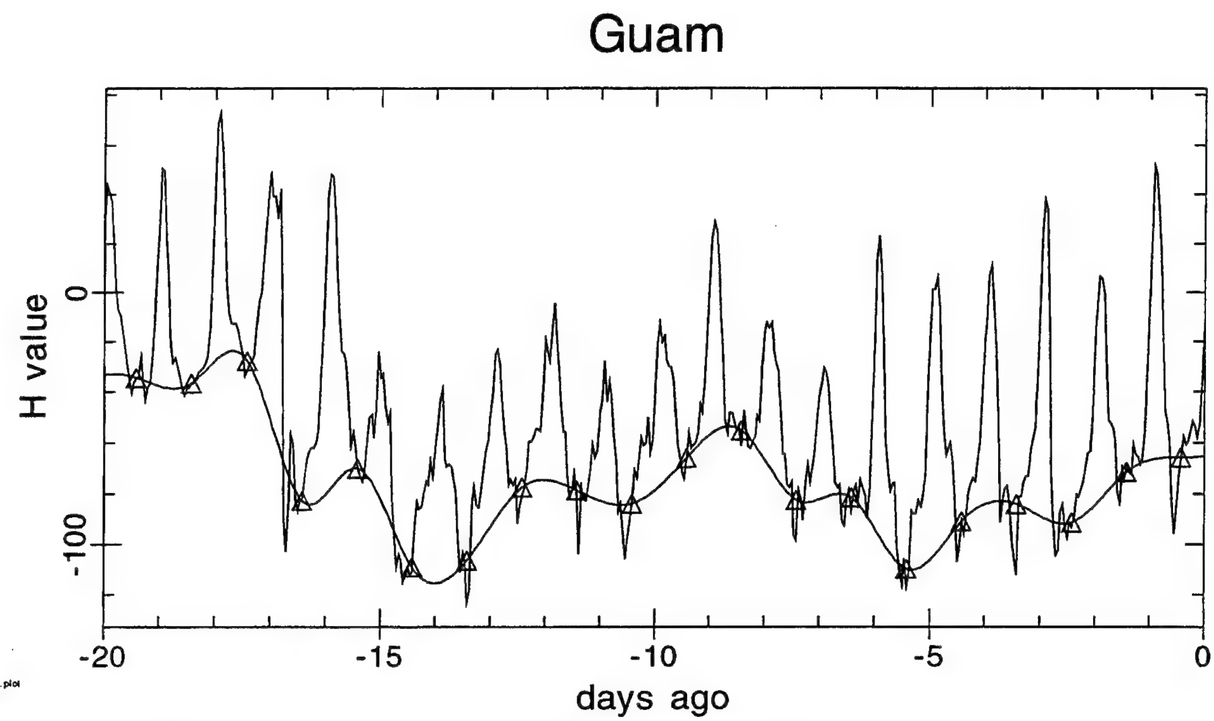


Figure 2. Cubic spline fit to the H values near midnight to "de-storm" the Guam magnetometer data.

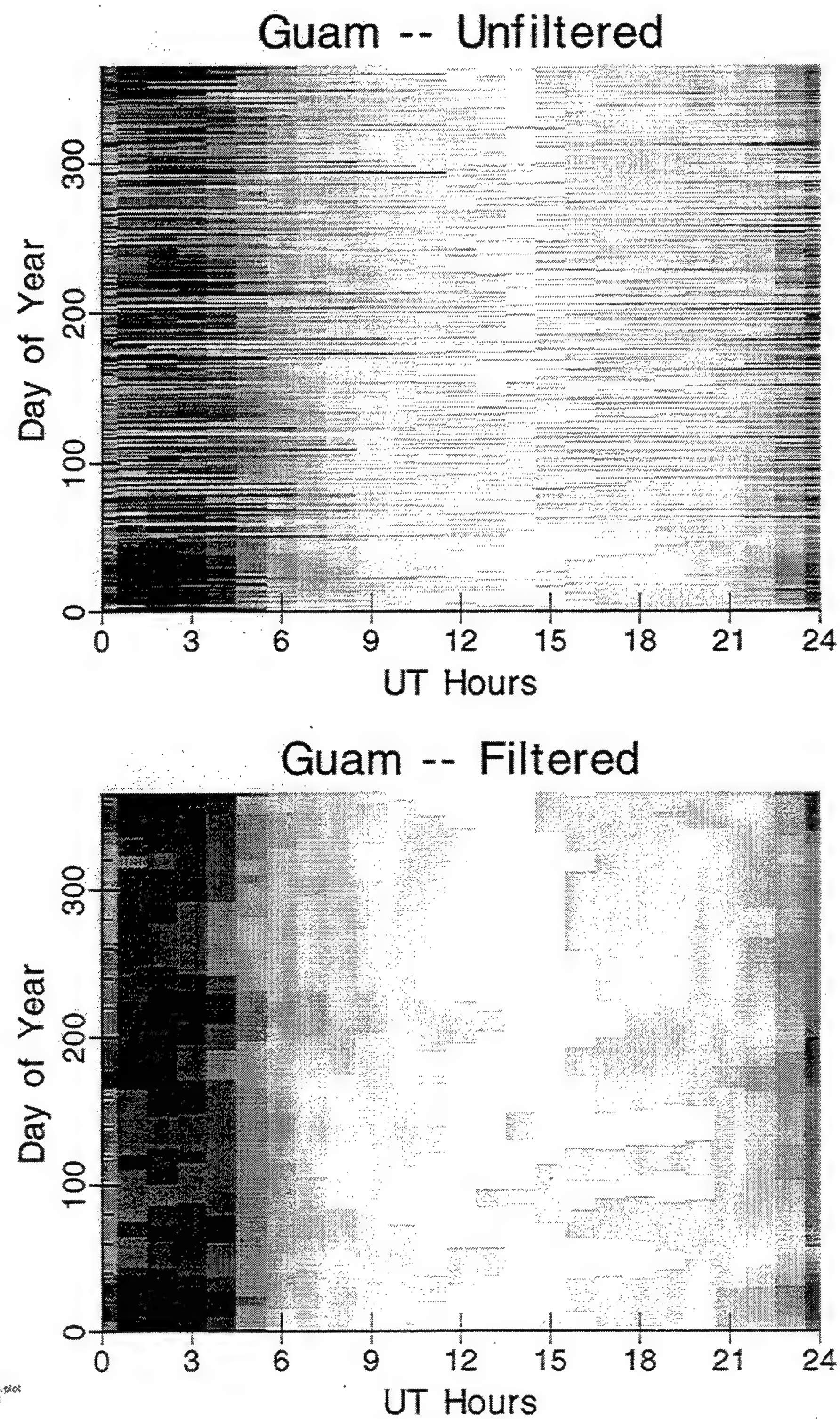


Figure 3. Unfiltered and Filtered data organized in two-dimensions of day-of-year and UT hours. The FFT filter is applied only in the day-of-year direction.

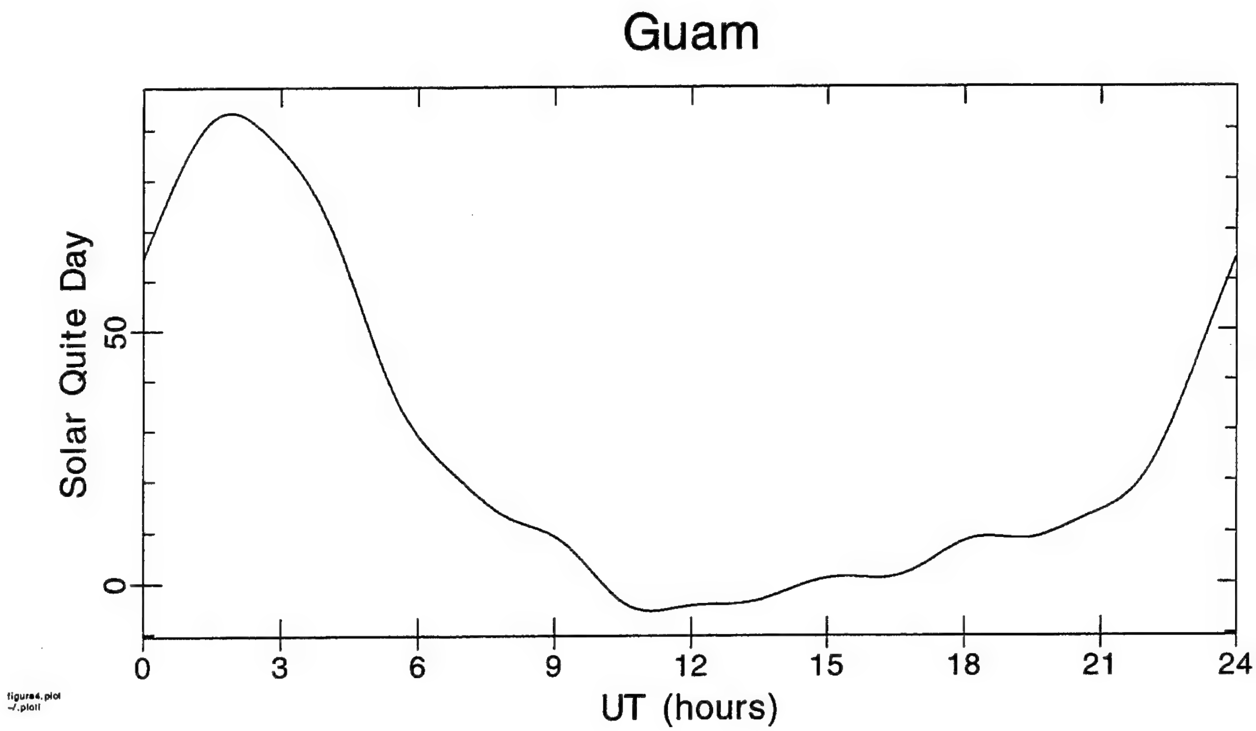


Figure 4. The Solar Quiet (SQ) variation of the H component from the Guam magnetometer station.

3.4. Dst from ΔH through Regression Coefficients.

The Dst approximation can be obtained from a single station measurement of the H value by removing the secular variation and the quiet day variation within the H measurements (equation 1) then relating the remainder to Dst at each UT (equation 3). The A_s and B_s coefficients (Figure 5) are obtained by performing a linear regression on a historical database of Dst and the ΔH's from each station measurements.

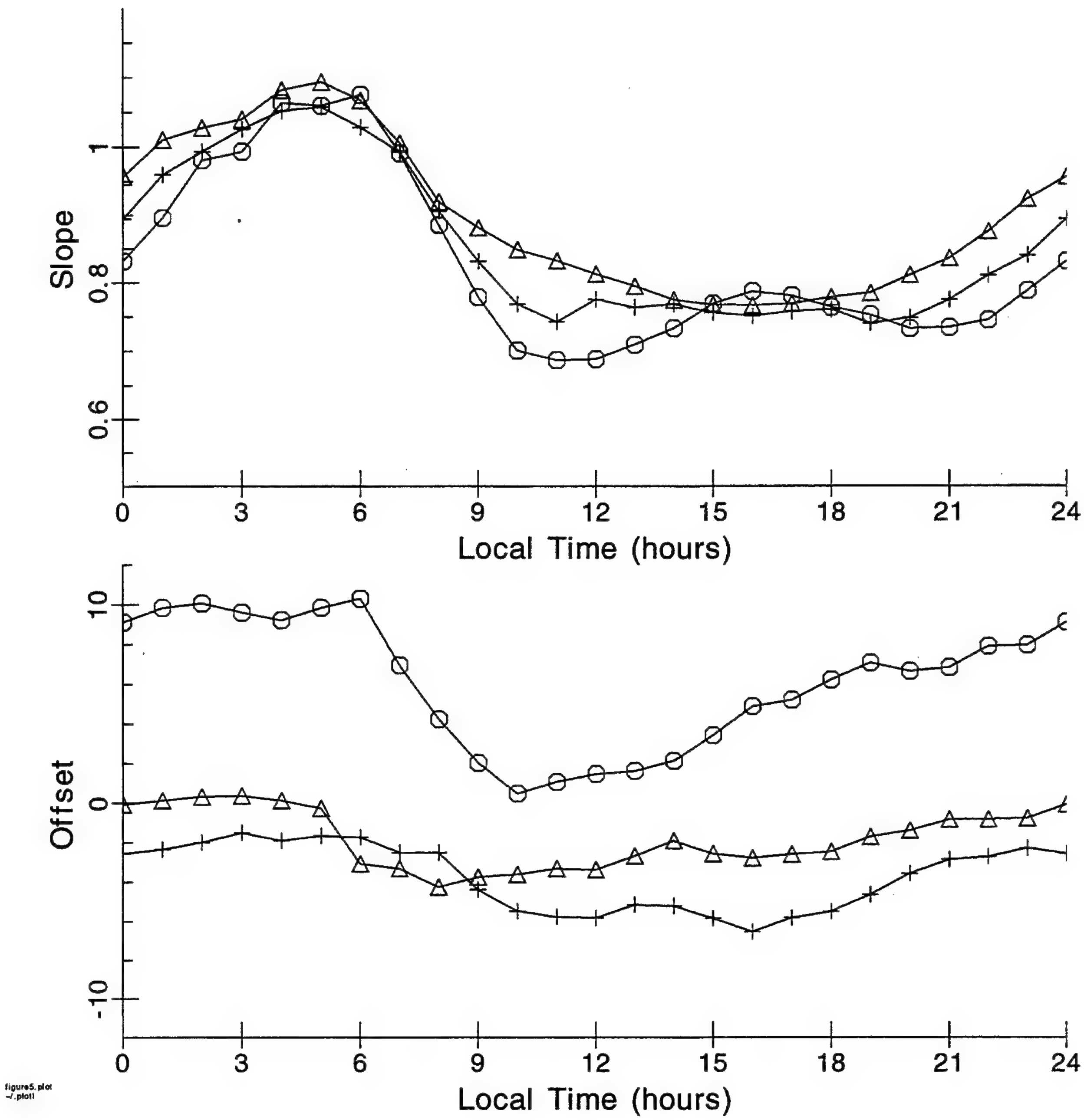


Figure 5. Slope and offset coefficients of the linear regression mapping of DH to Dst for San Juan (circles), Honolulu (triangles), and Guam (plusses).

3.5. Single RDst Approximation

Each station is used to estimate the real Dst through the above algorithms. A single best estimate to Dst is provided by averaging all available station estimates.

4. PROGRAM STRUCTURE

4.1. Introduction.

The main program of DST_CALC contains the modular programming structure desired in the FORTRAN standards documentation, *FORTRAN-77 Computer Program Structure and Internal Documentation Standards for the Air Force Forecast Center*. The main program calls, first, input modules (READ_IN_FILES & UPDATE_DBS & IN_TRENDS), second, calculation modules (DO_TRENDS & DO_DST & DO_DST_ERROR), and, finally, the output module (WRITE_OUT_FILES). There is an additional module, DO_CLEAN_MINUTELY inserted in the input phase which cleans the input data. Each module contains most of the subroutines required with the exception of public domain software of numerical methods, date manipulation, and character manipulation. UPDATE_DBS is not strictly an input module but a module for maintaining the necessary database files read in the READ_TRENDS module. Figure 6 diagrams the body of the main program DST_CALC.

DST_CALC runs silently unless an error is encountered. Errors observed in the program are either mitigated within the program silently or reported using the Space Forecast Center computer messaging utility.

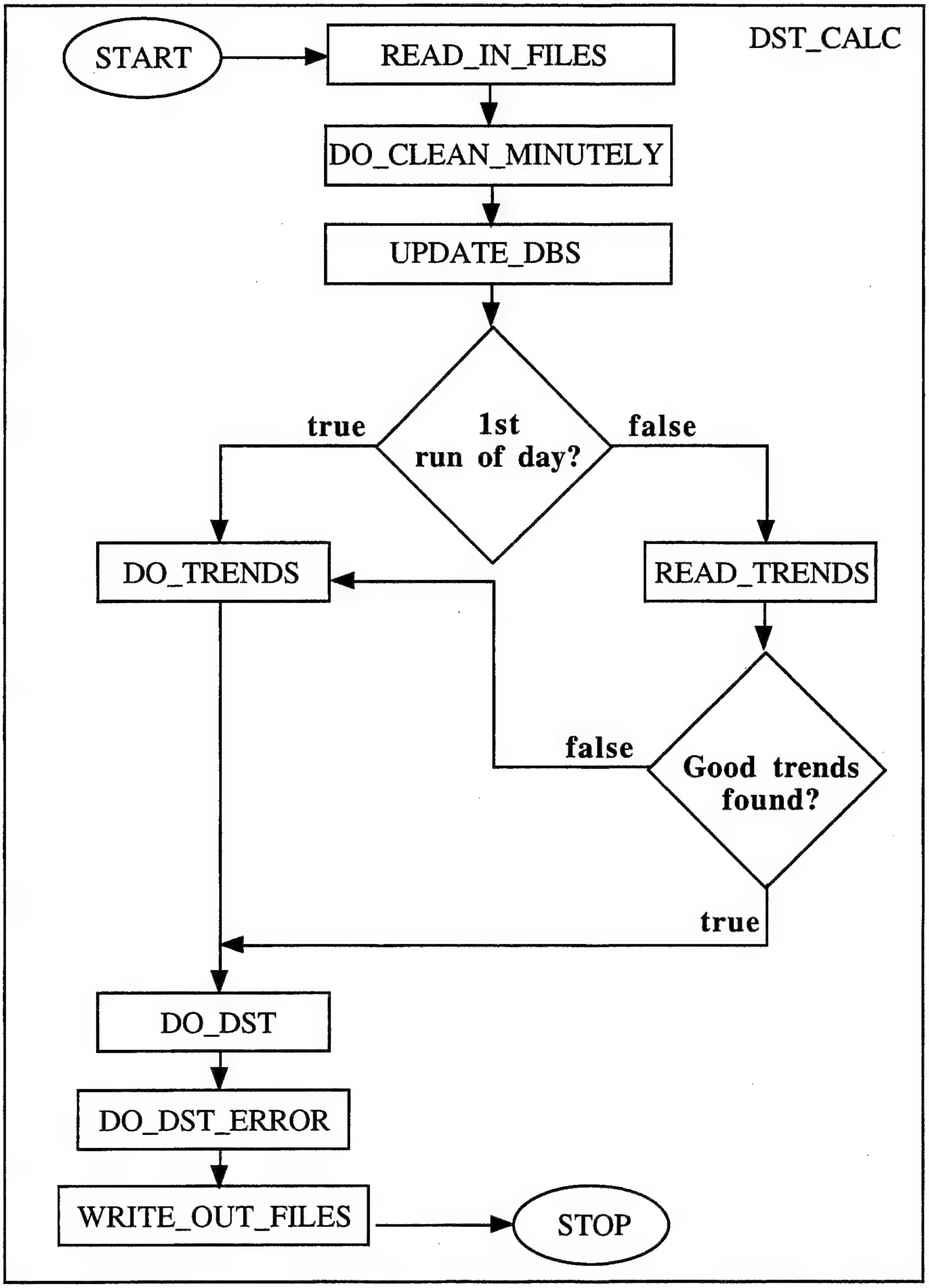


Figure 6. Diagram for the main program modules in DST_CALC.

4.2. Input Modules.

4.2.1 READ_IN_FILES Module. There are several data and message files created by the DST.COM controller program that must exist before DST_CALC is run. The two subroutines in the READ_IN_FILES module are READ_MSG_FILES, which reads message files, and READ_MAG_DAT, which reads the data file containing the H values of the magnetometer stations. Figure 7 diagrams the module and lists the files read. READ_IN_FILES also determines whether or not the present run is the first run of the day.

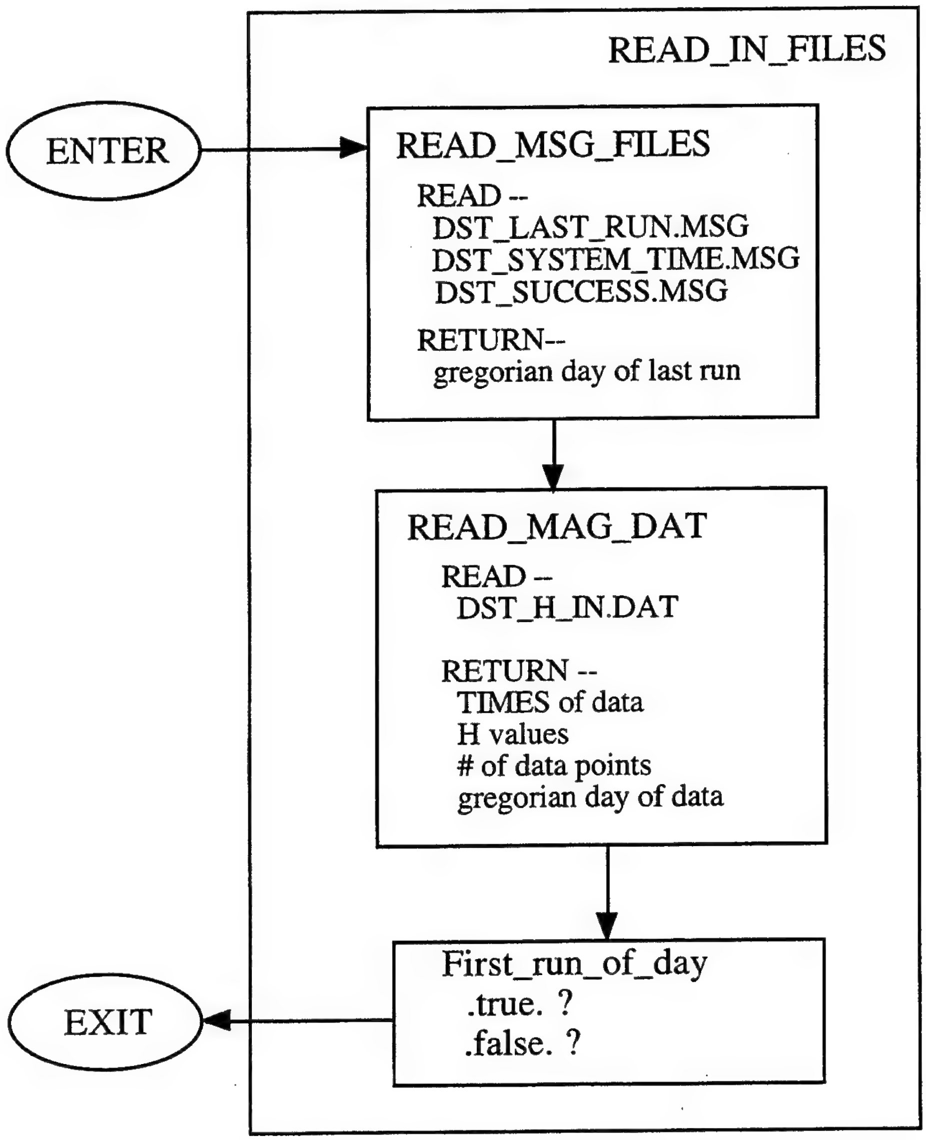


Figure 7. Diagram for the READ_IN_FILES module.

4.2.2. Data Cleaning Module. The data input into DST_CALC may have problems of spikes, offsets, and gaps. A module was placed into the DST_CALC program to removed spikes. This was a simple addition due to the readily available 'spike' remover available at Space forecast Center. It is a robust and simple algorithm that removes all 'stand alone' excursions in the data. It is a heavy-handed removal but will not damage the DST approximation. Of more concern is the damage to DST approximation by data problems. This is a place holder subroutine for future data cleaning algorithms that are determined necessary for maintaining quality DST approximations.

4.2.3 UPDATE DBS Module. DAT_CALC requires a knowledge of the historical horizontal component, H, for each station. This historical database must be updated on an ongoing basis. The historical information is maintained by the UPDATE_DBS module in 4 files associated with each magnetometer station and 1 file containing a ten year history of Kp.

```
DST_CALC_TODAYxxxxxx.DAT
DST_CALC_YESTERDAYxxxxxx.DAT
DST_CALC_HOURLYxxxxxx.DAT
DST_CALC_DAILYxxxxxx.DAT
DST_CALC_KP.DAT
```

These files must be present when DST_CALC is run. Termination error will result if any one is absent. The files are updated by DST_CALC, but the cannot be created DST_CALC. The data within the files can contain BAD data flag of 99999.9. The 'xxxxxx' of the above file names is a place holder for the station number or WMO number of each station being used in the DST approximation. In the initial configuration of DST_CALC and DST.COM there will be three stations and, thus, three files each of the TODAY, YESTERDAY, HOURLY, and DAILY files.

DST_CALC_TODAYxxxxxx.DAT files contain the present day's minutely values of the horizontal component, H, of the magnetic field at the xxxxxx station. These TODAY files are updated with every run of DST_CALC using the H values obtained from the DST_H_IN.DAT file. Future times of the present day in the H_TODAY array contain 99999.9 data flags. The files are unformatted files created with the following write statements:

```
WRITE(LIN)I_GREGORIAN_TODAY
WRITE(LIN)(H_TODAY(I),I=1,1440)
```

I_GREGORIAN_TODAY is an INTEGER*4 variable containing the Gregorian

day count from day 1 of year 1 AD of the present day. H_TODAY is a REAL*8 array containing the minutely values H of each station.

DST_CALC_YESTERDAYxxxxx.DAT files contain the previous day's minutely H values. These files are updated by DST_CALC when a new day is detected. On the new day, H_TODAY data becomes the H_YESTERDAY data and H_TODAY is filled with 99999.9's. The DST_CALC_YESTERDAYxxxxx.DAT files are unformatted files and created with the following write statements:

```
WRITE(LIN)I_GREGORIAN_YESTERDAY
WRITE(LIN)(H_YESTERDAY(I),I=1,1440)
```

I_GREGORIAN_YESTERDAY is an INTEGER*4 variable containing the Gregorian day count from day 1 of year 1 AD of the previous day. H_YESTERDAY is a REAL*8 array containing the minutely values of H for each station.

DST_CALC_HOURLYxxxxx.DAT files contain the hourly averages of H for 365*24 hours with the last array element holding the hourly average H of yesterday's 23rd UT hour. These files are updated by DST_CALC when a new day is detected. On the new day, the hourly average H data is shifted within the array H_HOURLY array by one or more days then new 'yesterday' hourly averages are calculated from the newly updated H_YESTERDAY data. The DST_CALC_HOURLYxxxxx.DAT files are unformatted files and created with the following write statements:

```
WRITE(LIN)I_GREGORIAN_YESTERDAY
WRITE(LIN)(H_HOURLY(I),I=1,365*24)
```

I_GREGORIAN_YESTERDAY is an INTEGER*4 variable containing the Gregorian day count from day 1 of year 1 AD of the previous day. H_HOURLY is a REAL*8 array containing the hourly averages H for each station.

DST_CALC_DAILYxxxxx.DAT files contain the daily averages of H for 365*10 days with yesterday's daily average as the last day of the data array. These files are updated by DST_CALC when a new day is detected. On the new day, the daily averages are shifted within the holding array, H_DAILY, then newly updated H_HOURLY data of 'yesterday' is averaged to obtain yesterday's daily average H. The DST_CALC_DAILYxxxxx.DAT files are unformatted files and created with the following write statements:

```
WRITE(LIN)I_GREGORIAN_YESTERDAY
```

```
WRITE(LIN)(H_DAILY(I),I=1,365*10)
```

I_GREGORIAN_YESTERDAY is an INTEGER*4 variable containing the Gregorian day count from day 1 of year 1 AD of the previous day. H_DAILY is a REAL*8 array containing the daily averages H for each station.

The DST_CALC_KP.DAT file contains the 366 days of hourly Kp values. The hourly Kp's are provided by SFC. The last 24 elements of the Kp array contain today's values of Kp. The file is updated on the first run of a new UT day. The Kp's are obtained from the DST_KPF10_IN.DAT file. The DST_CALC_KP.DAT file is an unformatted file and is created with the following write statements:

```
WRITE(LIN)I_GREGORIAN_TODAY
WRITE(LIN)(XKP(I),I=1,366*10)
```

I_GREGORIAN_YESTERDAY is an INTEGER*4 variable containing the Gregorian day count from day 1 of year 1 AD of the previous day. XKP is a REAL*8 array containing the hourly Kp values.

The subroutines within the UPDATE_DBS module are:

MINUTELY_UPDATE	- updates H_TODAY every run
PRESENT_TO_PREVIOUS	- updates H_TODAY & H_YESTERDAY
HOURLY_UPDATE	- updates H_HOURLY
DAILY_UPDATE	- updates H_DAILY
KPF10_UPDATE	- updates Kp

Figure 8 diagrams the flow of the UPDATE_DBS module.

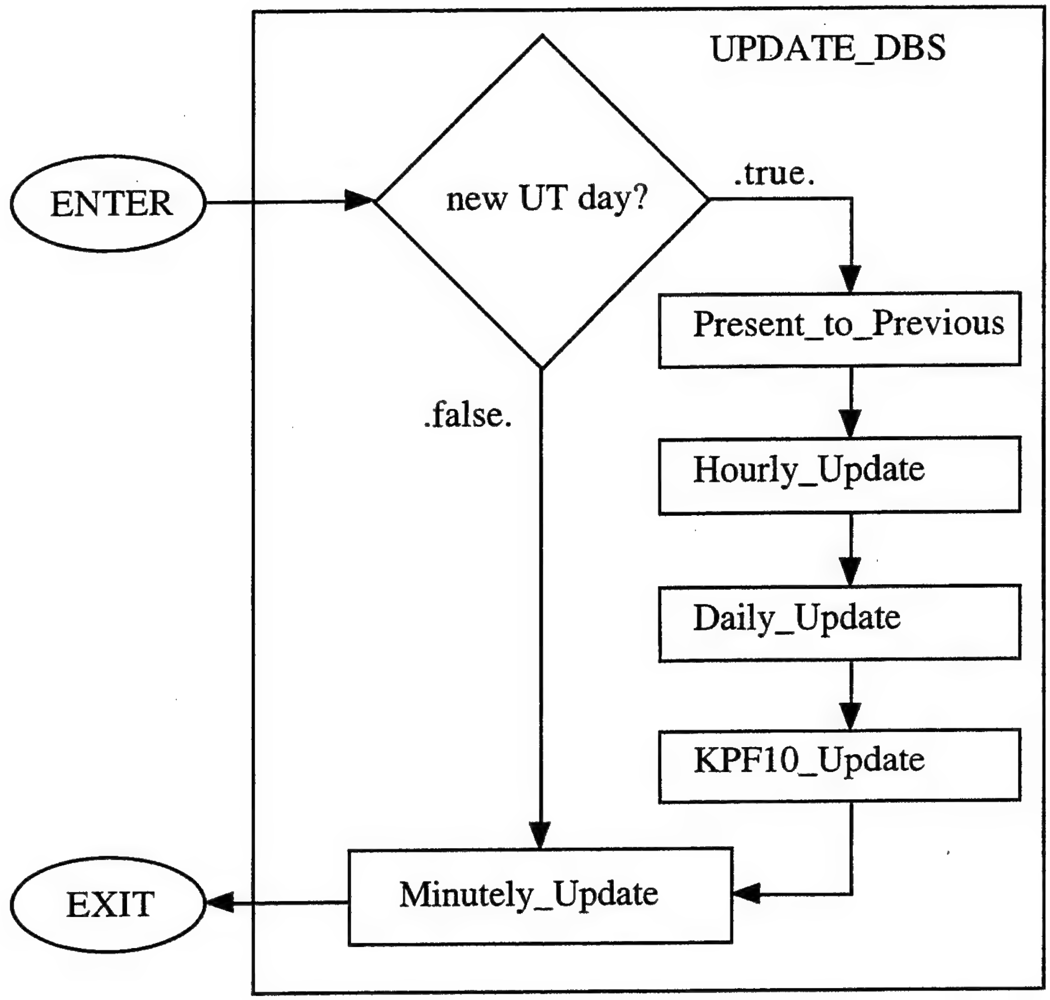


Figure 8. Diagram for the UPDATE_DB module.

4.2.3 READ_TRENDS Module. DST_CALC creates trend files based on the historical database of hourly and daily averages of H. These trends --- Secular Variation, Quiet Solar Variation, and UT Regression Coefficients --- are created on the first run of a new UT day and output DST_CALC_HSV.DAT, DST_CALC_HSQ.DAT, and DST_CALC_HCO.DAT, respectively. Every other call in the UT day, DST_CALC will read the trend files in the module READ_TRENDS. READ_TRENDS makes several function calls --- IN_SECULAR, IN_QUIETDAY, IN_COEFF. Each subroutine/function reads a trend file. If the files are read successfully a logical flag GOOD_TRENDS is set to true. READ_TRENDS returns the logical flag value to the calling program. If the trends are not read successfully then DST_CALC will try to recreate the trends. The READ_TRENDS module is diagrammed in Figure 9.

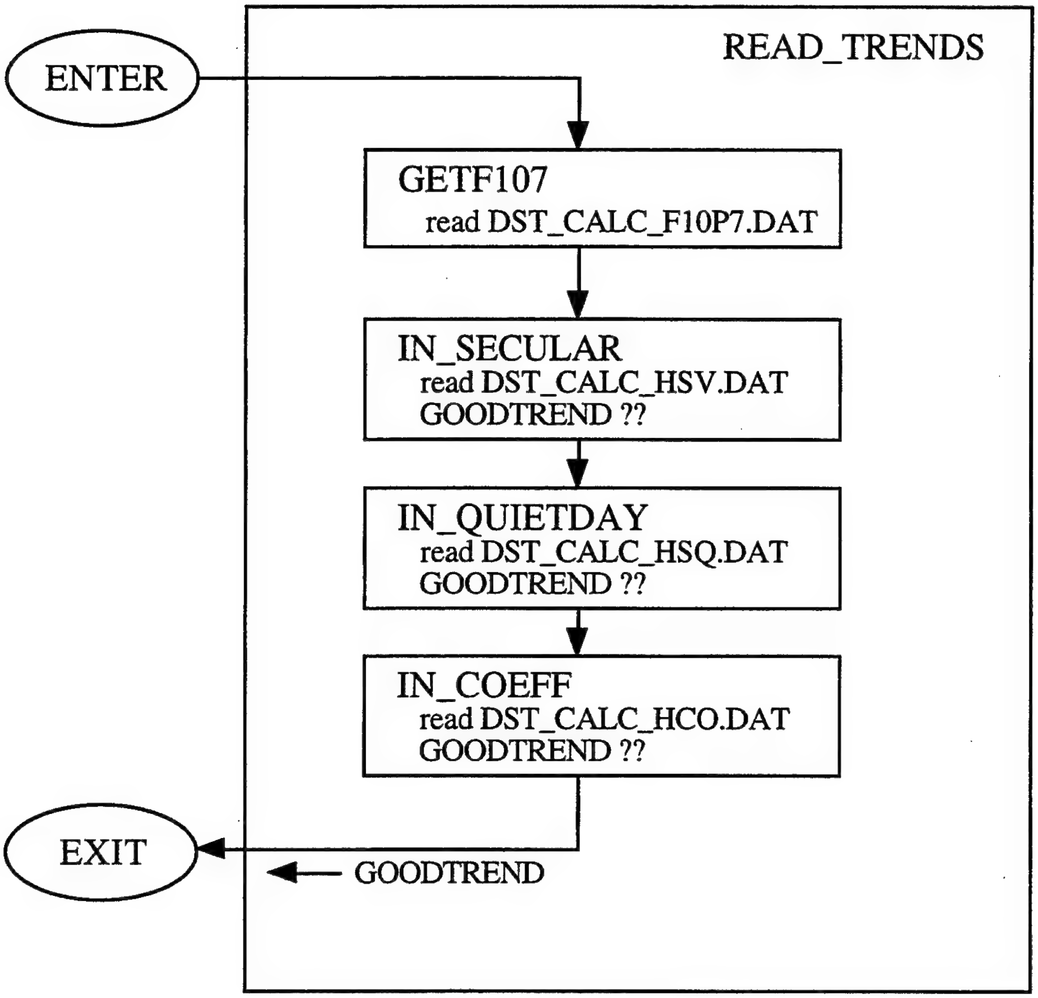


Figure 9. Diagram for the READ_TRENDS module.

There are four files read in READ_FILES module. The first file read is DST_CALC_F10P7.DAT, which contains the most recent 10.7 cm solar flux value as provided by DST_COM (DST_KPF10_IN.DAT). DST_CALC_F10P7.DAT is a formatted file written using a free format:

```
WRITE(LUN,*)IF10P7
```

The value of IF10P7 can vary from 60 to 400 and has a BAD flag of 9999. If there is an error in the file or the value of IF10P7, then the variable is set to a medium solar value of 150 and the program continues. This is not critical information of the operation of the program.

The secular variation does not change in the course of a day so the

DST_CALC_HSV.DAT file contains a single value of HSV for each station. If HSV was incalculable then it will hold 99999.9. The file is a formatted file written using free formats:

```
WRITE(LUN,*)IGREGORIAN_TODAY
WRITE(LUN,*)(I_STATION(IS),IS=1,N_STATIONS)
WRITE(LUN,*)(HSV(IS),IS=1,N_STATIONS)
```

IGREGORIAN_TODAY is the Gregorian day integer. N_STATIONS is the number of stations possible in the current configuration of DST_CALC and DST_COM. N_STATIONS appears as an integer parameter in PARAM.INC. I_STATION is an array containing the WMO numbers for the separate stations.

The quiet day variation is defined for each minute of the current day and the file DST_CALC_HSQ.DAT contains the SQ variation for each minute for each station. If HSQ was incalculable then it will hold 99999.9 in all minutes of the day. The file is an unformatted file:

```
WRITE(LUN)IGREGORIAN_TODAY
WRITE(LUN)(I_STATION(IS),IS=1,N_STATIONS)
DO IM=0,1440
     WRITE(LUN)(HSQ(IM,IS),IS=1,N_STATIONS)
ENDDO
```

The regression relation, which relates the ΔH value to DST, is a linear relation for each minute of the UT day (equation 3). The file containing the 1440 minutely values of A and B, DST_CALC_HCO.DAT, is an unformatted file:

```
WRITE(LUN)IGREGORIAN_TODAY
WRITE(LUN)(I_STATION(IS),IS=1,N_STATIONS)
DO IM=0,1440
     WRITE(LUN)(HAA(IM,IS),HBB(IM,IS),IS=1,N_STATIONS)
ENDDO
```

HAA is the slope of the regression relation, HBB is the intercept of the regression relation.

<u>4.3. Calculation Modules.</u>

<u>4.3.1. DO_TRENDS Module.</u> The module DO_TRENDS contains the subroutines that calculate the secular variation of H, HSV, the solar quiet day of H, HSQ, and the regression coefficients of the Delta H-to-Dst relationship, HAA

and HBB. DO_TRENDS outputs the trends to the files DST_CALC_HSV.DAT, DST_CALC_HSQ.DAT, and DST_CALC_HCO.DAT, respectively. The basic structure of DO_TRENDS is simple and is diagrammed in Figure 10.

The secular variation of the horizontal component of the earth's magnetic field is the slow changing component of H caused by the earth's intrinsic field. Using TODAY as day zero, the daily average values from DST_CALC_DAILYxxxxxx.DAT are divided into 27-day periods back for nearly ten years. Each 27 day set of daily averages are sorted by average H value. The top 20% of the daily average H's of the 27 day set are averaged. These become the 80th-percentile averages of each 27 day period. Using the 80th-percentile averages removes days when storms depress the daily average. Using the collection of the 27-day 80th-percentile averages DO_SECULAR calls a Least Square Fit routine to obtain a best fit 5th order polynomial to define HSV.

The solar quiet variation of the horizontal component is the daily repeatable portion of the H component. The hourly averages of the previous 365 days obtained from the DST_CALC_HOURLYxxxxxx.DAT files. To obtain the quiet day one must remove, first, the secular variation, second, storm influences, third, extract the smoothly-repeatable portion of H. The secular variation is removed from the year of H hourly averages by subtracting the best-fit 5th-order polynomial, HSV, found in DO_SECULAR. To remove storm influences a cubic spline fit to local midnight values of H-HSV is created and subtracted from H-HSV. Finally, possible sub-storm influences are removed by placing BAD flags in the 'destormed' values during times of high Kp, 4.0 or larger. There is still considerable day-to-day variation, so smoothing and filtering will be used to obtain a best quiet day variation for each station. Unfortunately, there are large and small gaps in the data which must be filled before FFT filtering can be applied. Small gaps (less than four hours) are filled with a linear interpolation between valid points. Large gaps are filled by obtaining a 14-day average of valid 'destormed' data at each UT hour before and after the gap. This produces an average quiet day before and after the gap. The gap is then filled by interpolating between these two average quiet days. There are no longer any gaps unless there was no data at all in the DST_CALC_HOURLYxxxxxx.DAT file.

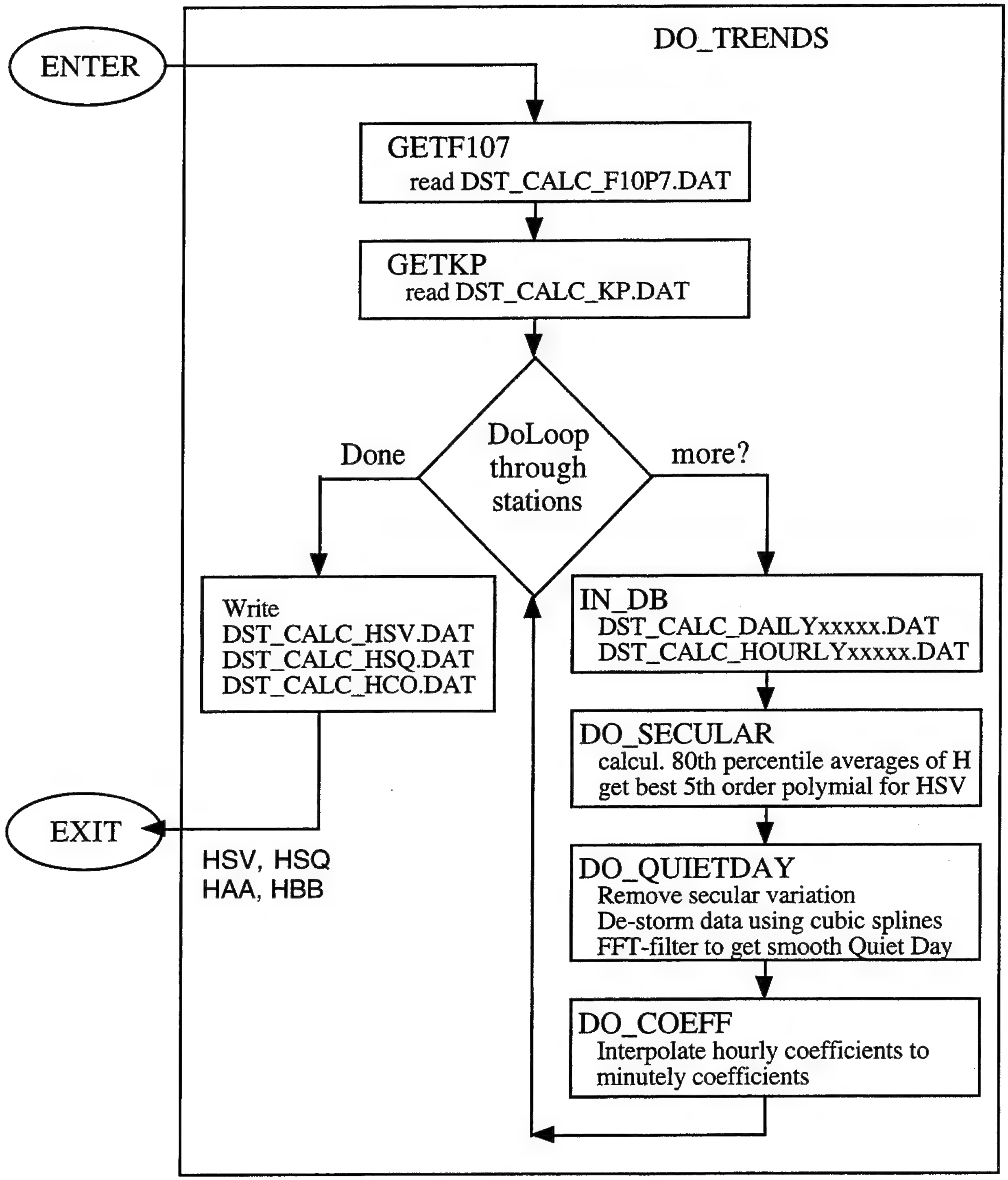


Figure 10. Diagram for the DO_TRENDS module.

Now we move to the last step for obtaining TODAY's quiet day variation. The de-stormed hourly averages are placed in a two-dimensional array with dimensions of UT hours and 'days-ago'. There is a zero day which represents TODAY. This is actually filled with data from the 365 'days-ago' day to produce a periodic boundary condition for FFT filtering. An FFT filter is used to remove the high frequency component in frequency space. The one-dimensional filter is applied in the dimension of 'days-ago' at each UT hour separately. Once each UT slice is filtered, then the zero day (TODAY) is smoothed by obtaining the first 10 Fourier coefficients of the zero day and using the coefficients to define the zero day solar quiet variation.

The regression coefficients relating of Eq (3) are based on historical data and are contained in a DST_CALC data statement (see, COEFF.INC). Robert McPherron of Space Environment Corporation performed the regression studies and provided the coefficients for each UT hour. The subroutine DO_COEFF interpolates the coefficients to provide minutely coefficients.

4.3.2. DO DST Module. Once the H value trends have been read in or calculated, the actual approximation of Dst is simple. DO_DST performs the calculations. The HSV_s and HSV_s are removed from the H_s values (Eq (1)), then the latitudinal dependence is removed via Eq (2), and, finally, the regression coefficients are applied via Eq (3). If H_s, HSV_s or HSQ_s have BAD flags (99999.9) because data contained bad flags or HSV_s or HSQ_s could not be determined, then Dst_s will be given BAD flag value (99999.9). To obtain the single Dst approximation product, the valid Dst_s are averaged.

4.3.3 DO DST ERROR. A root-mean-square (RMS) error estimate is provided to the Dst approximation. The subroutine DO_DST_ERROR provides the RMS error estimate though a simple look-up table of errors created from RMS studies performed on the RDST approximations using the above described single station algorithms on historical data. The RMS studies were performed by Geoff McHarg of the Air Force Academy in Colorado Springs.

4.4. Output Module.

4.4.1. WRITE OUT FILES. The results of the above calculations are output by WRITE_OUT_FILES. The outputs include DST_OUT.DAT file, which contains the RDST approximation and error for every minute passed in via the DST_H_IN.DAT file. Additionally, the present day secular variation values for each station are written to DST_HSV_OUT.DAT and the present day quiet day variation values for each station are written to DST_HSQ_OUT.DAT. These output files are discussed in the RADEX SVS documentation.

5. PUBLIC DOMAIN SOFTWARE USED IN DST_CALC

5.1. NETLIB

The National Science Foundation funds a world wide web site dedicated to numerical methods developed under NSF funding. The contents are public domain codes written in C and FORTRAN. Typically, the requirement for use is a note of acknowledgement to the code author with a description of how the code is used. Note that the codes are of industrial quality. Much of the 'Numerical Recipes' algorithms are based on the codes contained in NETLIB libraries.

5.1.1. FFTPACK. FFTPACK contains subroutines useful in FFT analysis and filtering of data. DST_CALC uses the forward and reverse FFT subroutines as well as a component analysis subroutine. For the DST_CALC program the FFTPACK subroutines were modified to use DOUBLE PRECISION real variables throughout.

5.1.2. NAPACK. The best fit polynomial routines within DST_CALC depend on a Singular Value Decomposition algorithm found in the NAPACK library of NETLIB. The subroutines were changed to use DOUBLE PRECISION reals throughout.

5.2. JVE UTILITIES

5.2.1. mynaJVE.f. There are several numerical subroutines written by J. Vincent Eccles available to the public without cost, expectation, or guarantee. The subroutines used in DST_CALC are:

MYSORT which sorts data in magnitude,
MYMEDIAN which determines a median value of a data set,
MYCUBIC which performs a cubic spline fit to data,
LEASTSQUAREFIT which fits an nth order polynomial fit to data.

5.2.2. dateJVE.f. There are several subroutines dedicated to manipulation of dates and years -- subroutines that calculate leap years, day-of-year values from dates, and Gregorian day values from dates, etc.. These were written by J. Vincent Eccles and are available to the public without cost, expectation, or guarantee.

5.2.3. charJVE.f. There are several subroutines dedicated to character manipulation. These were written by J. Vincent Eccles and are available to the public without cost, expectation, or guarantee.

6. ADDING OR REMOVING OR CHANGING A STATION IN DST_CALC

DST_CALC is written to expect a particular set of stations. The data from each station must be present in the input file, though every data point can contain BAD flags. If a station is removed or added or changed, then DST_CALC FORTRAN must be edited and recompiled. The portions of DST_CALC that must be changed to accommodate station changes is limited to INCLUDE files and one subroutine.

The first include file to be changed is PARAM.INC. The procedure for altering the include file is documented in the PARAM.INC.

In PARAM.INC one must:

1. change variable N_STATIONS to the new number of station.
2. provide station number I#_STATNUMB for any new station.
3. provide geo-latitude and geo-longitude of the new station.

In COEFF.INC one must alter the data statements for both coefficients, HA (slope) and HB (offset) to include new station regression coefficients matching the real Dst to the ΔH of the new station. The coefficients are determined off-line by comparing historical Dst and a Dst approximate obtained from historical H data from the new station.

In subroutine STATION_SWITCH new statements must be added to account for any new stations. The subroutine is documented internally to guide the alterations.

Finally, the historical databases required to run DST_CALC must be created, that is, DAT_CALC_DAILYxxxxxx.DAT DST_CALC_HOURLYxxxxxx.DAT, DST_CALC_YESTERDAYxxxxxx.DAT, and DST_CALC_TODAYxxxxxx.DAT. These files are discussed in earlier sections. The yesterday and today files can be filled with BAD flags (99999.9) to begin DST_CALC. The xxxxxx represent the WMO number of the new station. This number must match the WMO number in PARAM.INC.

Conceptual Plan for Quality Management and Verification (QMV) of Data and Model Products Associated with the 50th Weather Wing's Space Models.

J. J. Sojka, V. Eccles, and R. W. Schunk

Space Environment Corporation
399 North Main, Suite 325
Logan, UT 84321

October 1995

Report produced as a result of SEC being tasked to address the question, how could quality control be added to an executive system without it growing arbitrarily to exceed a specific share of the computational CPU, DISK, and MEMORY resources?

Table of Contents

1. Introduction

2. Kp_QMV tutorial

3. Kp_QMV summary

4. Kp_QMV stand-alone options

1. INTRODUCTION

The executive system is not only responsible for the obvious day-to-day tasks of supporting the operators at 50th Weather Wing, but, it is also expected to track information relevant to the product quality. It is impractical to use operators to monitor the quality of data stored in data bases, new observations coming in on data links, or algorithm outputs being generated at 50th Weather Wing. Not only are there insufficient operators, they are not adequately trained to make quality assessments of the vast range of information at the 50th Weather Wing. Most of the products created at the 50th Weather Wing will involve a combination of new observations, archived information, empirical models, and physical models. The observations and archived data will define inputs for the model and after it has run, its outputs will form the basis of the product. The QMV problem is then a question first of all to ensure that data quality is maintained throughout such that inputs to model are uncorrupted; secondly, each model performance characteristics are monitored to determine the quality of their outputs; thirdly, since most of these individual data streams and models are complex and do not have a standard quality, the QMV must be able to archive information that will lead to such standard quality assessments; fourthly, the real-time operations supported by the executive system requires that the QMV be able to alert operators to poor quality situations.

Historically, QMV has not been viewed as an integral part of software product design in scientific sectors of space weather specification, i.e. empirical models IRI, MSIS or theoretical models TDIM, TGCM do not provide QMV information. Yet all four products are widely accepted and extensively used. In creating an almost autonomous executive system to be run by non-expert operators, the issue of QMV does not have a precedent in solar terrestrial space research. Hence, a tutorial example is first discussed to explore the potential tasks a QMV must carry out autonomously within the executive system. The example to be used is the Kp index, and it will be considered from two different but complementary points of view.

(a) SEC acquires Kp values over a network from SEL, NOAA

(b) 50th Weather Wing acquires magnetometer data that is then suitably scaled and processed to produce Kp values.

In both cases the final real-time Kp data stream is the most important time series input for the suite of space weather models currently under development for 50th Weather Wing. Its importance is two fold: it represents the key measure of

geomagnetic disturbance, and is the only space index which is readily and continually available. From the readers prospective it is one part of the entire process that probably is reasonably understood by all and, hence, is an ideal tutorial candidate. Contrary to expectations this simple index has been shown to be very difficult to implement in an operational environment. Hence, the tutorial is not without merit as a discussion of Kp quality.

The potentially prohibitive demands upon computational resources from the QMV going astray in the two scenarios is summarized in Section 3 and a well-behaved alternative QMV scheme is presented in Section 4.

Based upon this alternative method of handling QMV we consider the applications of such a technique to the ionosphere. In Section 5 the overall executive system is outlined from the ionospheric viewpoint, and key QMV issues are identified in Section 6. This latter section also outlines a QMV scheme for the ionosphere.

Applying the methodology one step further we address the overall executive system QMV problem in Section 7. Keeping clear the objectives and needs for which a QMV is designed. We summarize in Section 8 with recommendations for the implementation and utilization of QMV software and products.

2. Kp_QMV TUTORIAL

Two contrasting scenarios for the Kp*[1] index are explored as a conceptual exercise. The objective, in both cases, is to define procedures whereby the user (an operator, a model, or the executive system) will be provided an index of acceptable quality. Even the term acceptable will be explored.

2.1 SEC near real-time Kp

SEC like many other companies, Universities and Government laboratories is on the internet and able to acquire NOAA-SEL products. SEL provides the community a near real-time Kp index in the form of a regularly updated file that users can read. Hence, by simply opening up a network connection to SEL, looking for the appropriate file, and transferring it back to ones own computer one has the Kp index. *The naive assumption would be that this is an error free real-time process.* Lets assume it is not and that things can go wrong. Recall that

[1] Note for the purpose of this report the term Kp is used as a generic label for the 50th W.W. index, SEC index, and the Gottingen index.

the process must work without an operator's intervention and that irrespective of how it is obtained, a continuous stream of real-time Kp is needed.

Figure 1 is a conceptual outline of how SEC would implement QMV on the retrieval of the SEC Kp data. To successfully bring back a Kp in real-time, processes 1 through 4 in Figure 1 must be successful. Any one of them can fail. If a failure occurs it is still necessary to have a Kp value in the real-time database. Hence process 5, the SEC prediction of Kp, is not going to produce the same Kp as obtained from SEL. The QMV must create suitable quality flags as well as record for future reference that a problem has occurred.

This is where the complexity of the QMV procedure occurs. Considering in turn the failure modes indicated by processes 1 through 4 cover aspects of problems to be experienced by any automated algorithm. First, process 1 requires that a network link be established between two computers. This has nothing to do with Kp, but without it, it prevents Kp from being acquired. A QMV must be able to let operators know if this type of problem occurs frequently such that the appropriate network engineering can be undertaken to remove the problem. The second process requires that the SEL Kp generation program has successfully updated Kp. Although SEC will have no influence on the dependability of the SEL algorithm to generate new Kp's on time, it is necessary for the QMV to monitor this type of error such that corrective action can be carried out. This may actually be a matter of changing the times at which the SEC utility sequencer would launch the BRING_KP, i.e., delay the process by 15 minutes, this may be long enough for the SEL algorithm to complete 'with greater reliability' the generation of a new Kp index. The third process is a file transfer. Transmission errors can occur as well as disk full situations. Like the first process, this is independent of the Kp itself, but when an error occurs it will prevent real-time data arriving. Once process 4 is reached, the SEL Kp is available but the question of its quality needs to be determined. Threshold limits, spikes, and data drops are the simple things to check for. But off-line cross checking of the SEL and SEC's reproduction of Kp would also be an appropriate QMV activity.

The QMV report would only occur in the cases where process 5 is required to fill in missing data. Hence, under ideal conditions no QMV activity is needed. This last consideration raises the major concerns of a QMV system. The amount of resources used by the QMV itself depends upon the process being monitored, especially the departure from the nominal operation status. The amount of CPU time, disk storage space, and potentially the number of error messages being sent back to the executive system will under bad conditions grow. Somehow the

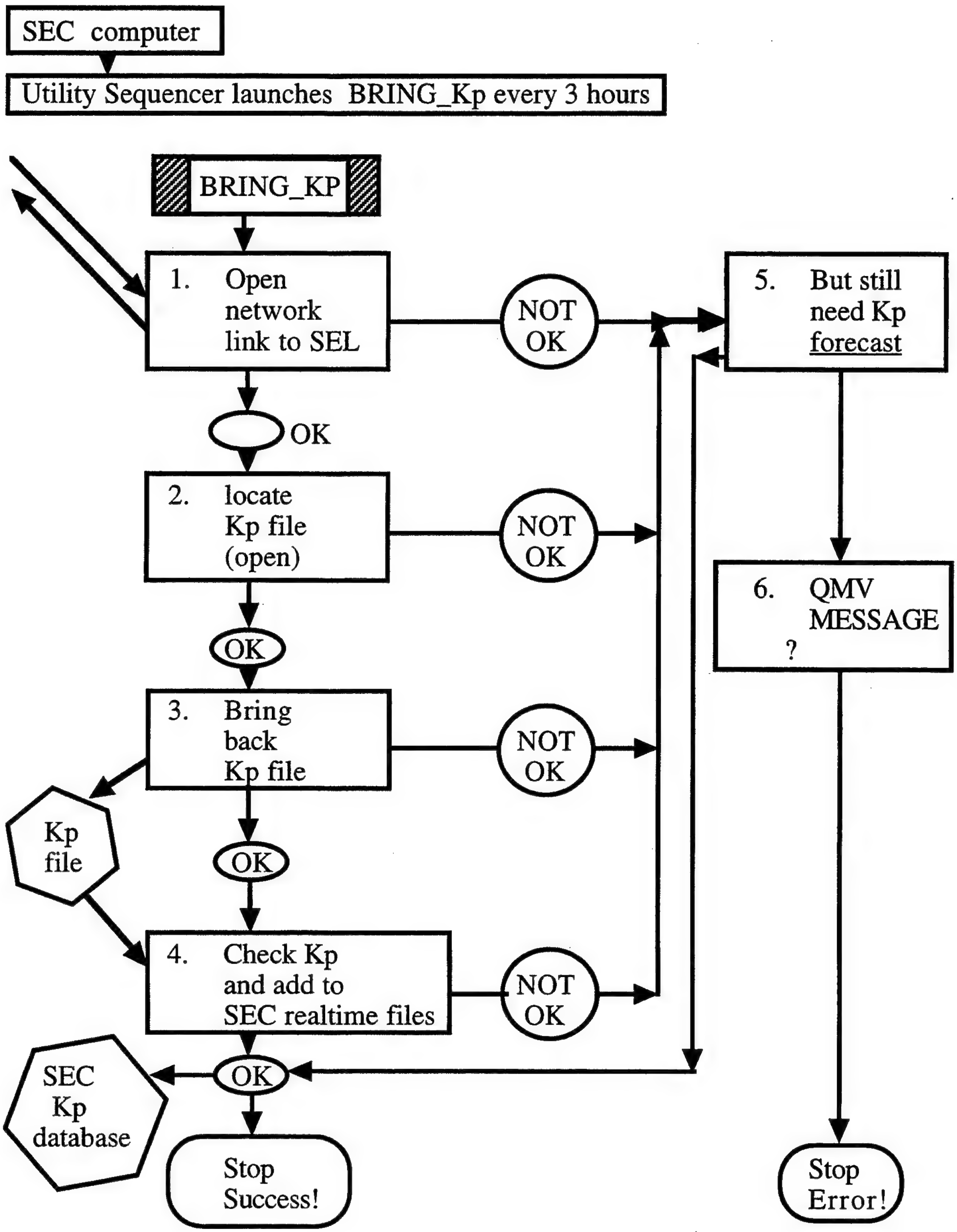

SEC computer
Utility Sequencer launches BRING_Kp every 3 hours
BRING_KP
1. Open network link to SEL
NOT OK
OK
2. locate Kp file (open)
NOT OK
OK
3. Bring back Kp file
NOT OK
Kp file
OK
4. Check Kp and add to SEC realtime files
NOT OK
SEC Kp database
OK
Stop Success!
5. But still need Kp forecast
6. QMV MESSAGE ?
Stop Error!

Figure 1

QMV-executive system must be configured to prevent the QMV hogging resources when situations are bad.

Consider again the Figure 1 scenario when BRING_KP operates correctly and data are good, one Kp value is stored every 3 hours and a data OK flag is set. If the data are bad and process 5 is invoked then a single Kp value will be forecast and an error flag set. But, in addition, process 6 is activated and information to record that an error occurred is generated. What kind of information is to be generated? This needs to be determined. Table 1 lists 5 levels of activity that process 6, the QMV, could carry out. Level 1, the least resource intensive, is simply to increment a single error counter. This counter could be monitored by the executive system, by the operators, or by regularly scheduled QMV reporting procedures. At this level no knowledge of the error scenario would be available other than by checking the quality flags of the Kp time series. The other extreme form is to record all relevant information at the time the error occurred and put it into a log file and simultaneously alert the executive system that an error occurred. This latter level is the type of QMV action one fears could take over the system. A regular occurrence of BRING_KP at the fifth level would quickly generate a log file significantly larger than the Kp data base. It would also use up more CPU time than the level 1 operation.

This topic will be discussed in more detail after the second Kp scenario is discussed.

Table 1. Complexity Levels of QMV Actions

1. Increment the error counter associated with the BRING_KP utility (a scalar).

2. Increment the appropriate error counter associated with the BRING_KP utility (a vector).

3. Increment the appropriate error counter, but also record a message that gives date and time information.

4. Increment the appropriate error counter, record the date and time information, but also include relevant information about BRING_KP status at the point of the problem.

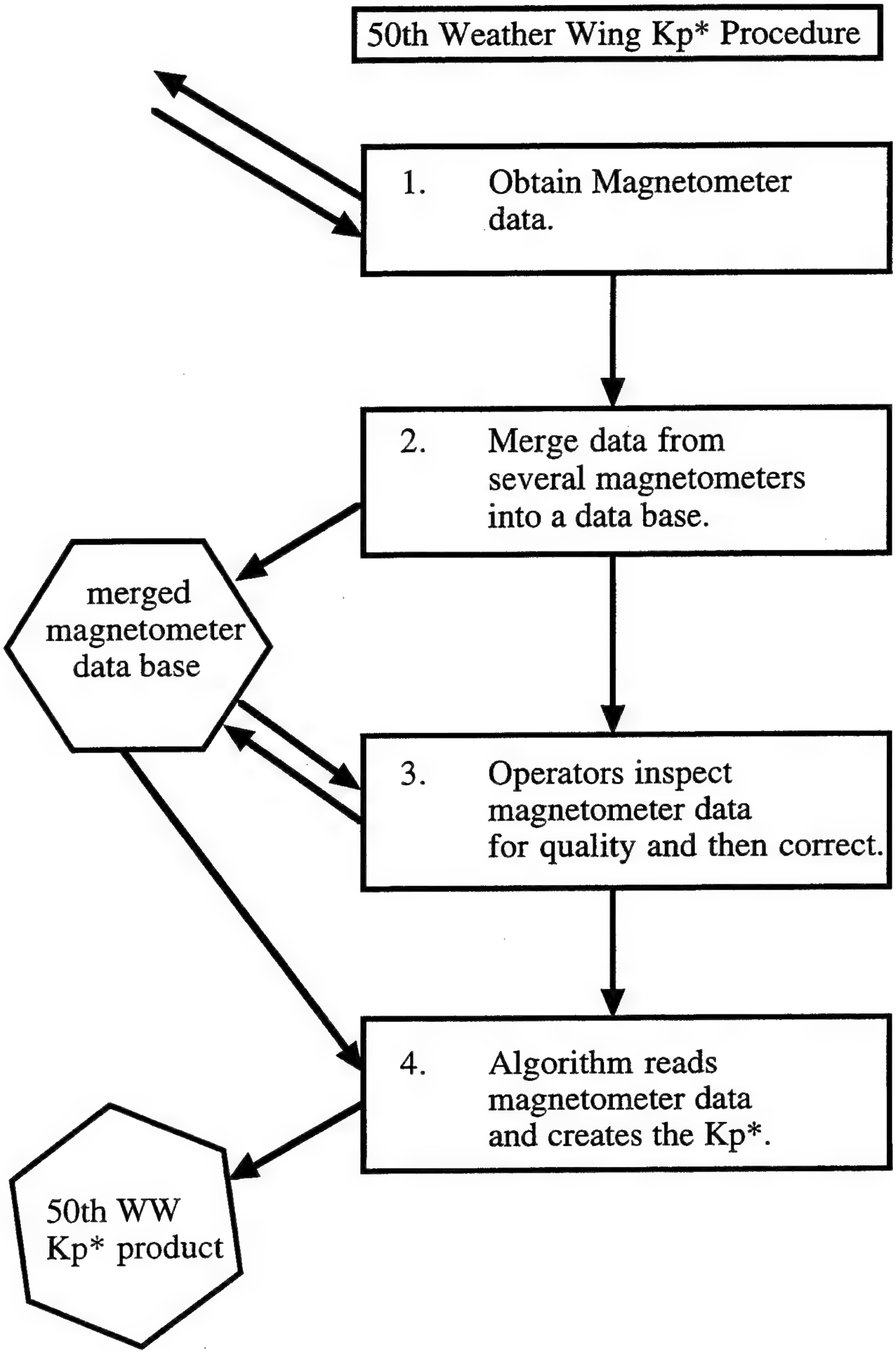
5. Same as 4, but also alert the executive system.

2.2 50th Weather Wing near real-time Kp

Data from ground-based magnetometers are transferred to the 50th W.W. in near-real time and operators carry out quality checks before an algorithm computes a Kp index from these data. Figure 2a summarizes the procedure for determining the Kp index. As more ground-based magnetometer data sets are obtained, the computed index becomes more like the Gottingen index. The Gottingen index is an internationally accepted standard and is called Kp. It is based on magnetometer data sets from a specific set of stations. Hence, any other combination will produce an index, but not necessarily one with the same characteristics as the Gottingen Kp index. Eventually, it is expected that the 50th W.W. will have enough magnetometer data that its Kp will be similar to the Gottingen Kp index. Since the algorithms that use the Kp index have all been developed using the Gottingen Kp index, it is crucial that an ongoing QMV effort establishes that indeed the two indices are equivalent.

The four procedures shown in Figure 2a are not supported by QMV software. Operationally, procedure 3 involves 50th W.W. staff inspecting the magnetometer data and when poor data are found they "fix" the data stream. At this time, when a 3 hourly or possibly 1 hourly index is being produced, the manpower to inspect these data is available. This would not be the case for a 15 minute type index. The use of operators to inspect and fix magnetometer data streams is not ideal. SEC has reviewed the magnetometer data stream quality issues and proposed automated procedures to carryout the Figure 2a procedure 3 work [SEC, USAF quarterly report Oct. 1994]. Specifically, Dr. McPherron presented details of how the Kp index was calculated and what data quality issues currently exist. Based upon these studies it is evident that procedures 1 and 2 are not without problems and, hence, also need to be monitored by a QMV such that the final Kp product can be suitably quality flagged.

Figure 2b is a revised version of Figure 2a in which hypothetical QMV operations have been incorporated as well as replacing procedure 3 operator intensive activity with an algorithm. Unlike the SEC near real-time Kp scenarios of the procedure section, the 50th W.W. Kp procedures are continually running and not necessarily coupled in the sequential manner indicated in Figures 2a and b. From the QMV activity procedures 1, 2, and 3 in Figure 2b can be viewed as a more complex form of the entire QMV activity outlined in Figure 1. The complexity arises from the fact that instead of a one index data stream, i.e., Kp in Figure 1, the 50th W.W. must acquire data from more than 12 magnetometer stations. Each one must be in near real-time, must be of good quality, and be merged such that the subsequent software will recognize data gaps, or missing



* the Kp index is not identical to the Gottingen Index.

Figure 2a

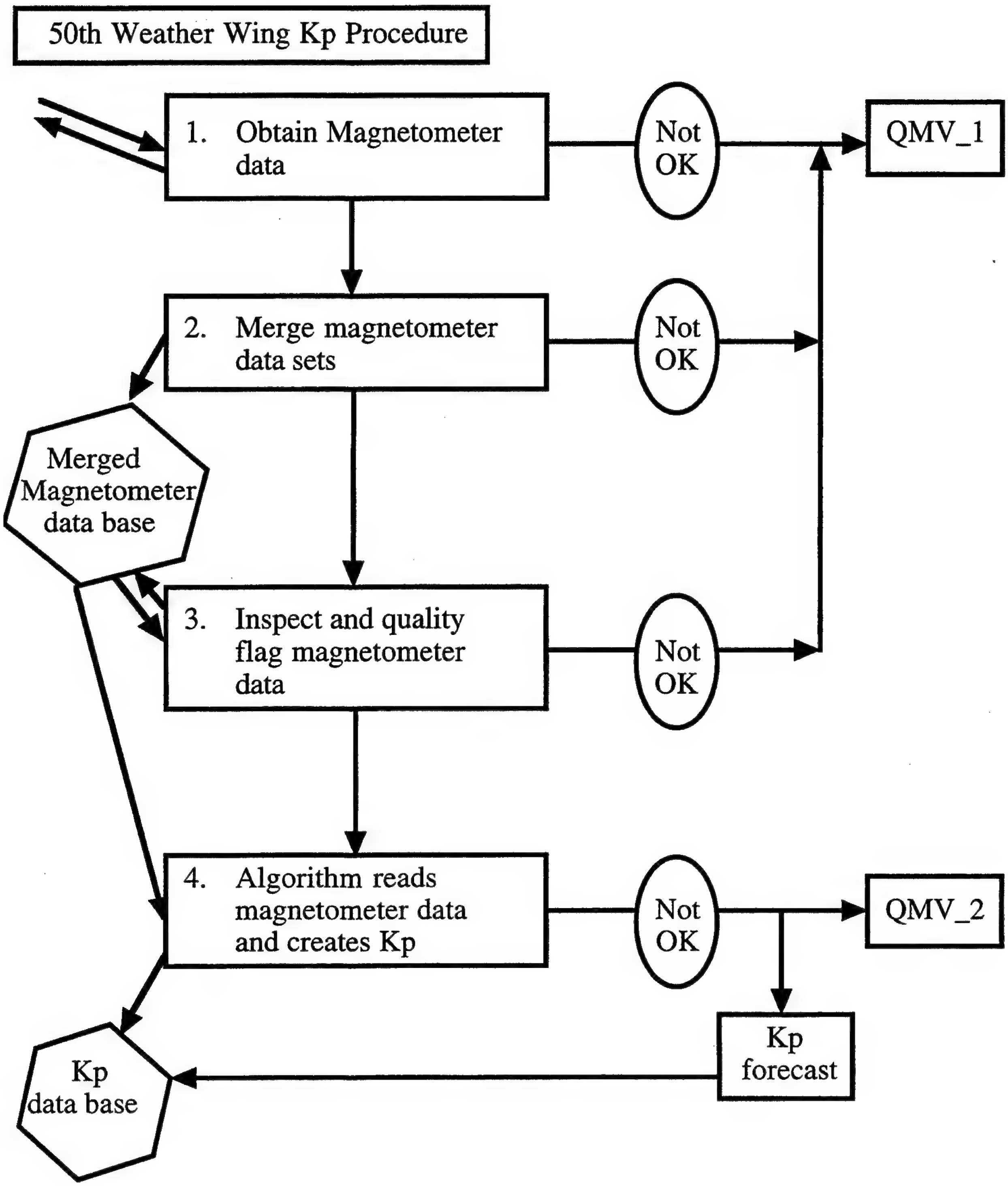

50th Weather Wing Kp Procedure
1. Obtain Magnetometer data
Not OK
QMV_1
2. Merge magnetometer data sets
Not OK
Merged Magnetometer data base
3. Inspect and quality flag magnetometer data
Not OK
4. Algorithm reads magnetometer data and creates Kp
Not OK
QMV_2
Kp forecast
Kp data base

Figure 2b

stations. Hence, the single QMV action identified in Figure 2b associated with the first three operations is somewhat more complex than the entire QMV activity shown in Figure 1. The exception shown in Figure 1 is that in the absence of a Kp a forecast of Kp had to be generated. This forecast is not needed in the first three procedures of Figure 2b. When data are missing and null values are substitute, it is crucial that all subsequent software recognizes the null values. The occurrence frequency of null values, their source, and bias towards specific stations would all be information to the QMV and follow-up improvement procedures. Hence, very much like the different options of QMV recording given in Table 1, these magnetometer data sets can be monitored at different levels of information generation.

The final procedure in Figure 2b is the calculation of Kp. This is a stand-alone algorithm that not only reads in the individual magnetometer data streams, but also a historic record of information that enables base-lines to be determined for each data stream. This entire process needs to generate QMV information, or in the event that insufficient data are available, flags of poor data quality must be set. In fact, since a Kp is always required, the algorithm must always produce a value. To ensure this happens, 4 may well have to include a Kp forecast capability used only in the event of no acceptable magnetometer data being available. At this extreme level, it would be reasonable for the QMV to alert the executive system, i.e. level 5 QMV activity in Table 1. Under these conditions, the quality of flagging associated with the Kp generation must adequately warn subsequent algorithms about the Kp status.

3. Kp_QMV SUMMARY

The preceding section develops an indepth QMV based on a detailed knowledge of how the BRING_KP utility operates. Thus, the development of the QMV must be done by the experts who developed the BRING_KP utility, or by someone who has become familiar with it. Many of the possible failure modes have nothing to do with the Kp itself, i.e., in both cases (Figure 1 and Figure 2b) the lack of communications link would result in no Kp.

Worse still in an operational environment where conditions can change, i.e. communication links can become noisy, the two BRING_KP schemes have the undesirable capability of generating more output as the error rate increases. An increase in output in all probability will also be reflected in an increase in QMV CPU time requirement.

A further difficulty of the outlined schemes is associated with potential

duplication of error reporting. Consider the case of a BRING_DST utility also operating. Then both the Dst and Kp utilities will have to have access to remote data via communication links. If the link is down, both utilities will independently identify, and report, this. Hence, a factor as much as 2 duplication in QMV activity will be generated.

Ideal and comprehensive as the outlined Kp QMV procedures are they are not what is needed. In fact SEC recommends that QMV based on the section 2 methodology is NOT developed. Table 2 summarizes the key reasons for not using this methodology.

Table 2. Reasons for NOT Building QMV Directly into Utilities.

- Requires expert knowledge of the operation of the utilities.
- Requires complete specification of input failure modes to the utility.
- QMV Software will always be buried inside the utility and not accessible. Hence, it must be extremely smart. This together with the first two items implies high cost and maintenance requirements.
- The amount of QMV will depend on the specific errors and their rates of occurrence. As the utility experiences increasing errors, its QMV activity will increase both in output reporting rates and CPU time.
- Duplication of error detection will occur, wasteful in resources.
- Much of the detailed QMV report will only be meaningful to the expert who developed the utility.

4. Kp_QMV STAND ALONE OPTION

Given that there are major difficulties and operation concerns about building QMV software into the utilities that generate Kp, how else can QMV be carried out? From a wider prospective, Kp is just one of many product generators that are running within the Space Models system. Each has been developed or obtained to supply a specified product. Hence, the other way to consider QMV would be to look at the output from the utility; in this instance the Kp generator. Kp is used by other utilities that have been developed on the internationally accepted Gottingen Kp index. Unfortunately, the Gottingen index

is not available in real-time nor is it available at this time within a secure environment. The Kp generator is therefore producing an index which is as close as possible to the Gottingen Kp.

On this basis the QMV activity can be broken into two parts. The first part would be the real time quality assessment of the Kp generators output. This output goes into a standard file and is time tagged. In the SEC scenario, Figure 1, new values of Kp are expected every 3 hours, while in the 50th W.W. scenario, Figure 2a, this generation of a new value would be as frequent as every 15 minutes. In both instances, the output data base is augmented at regular intervals. Hence, a quality assessment could be made at these regular intervals of the newly generated data. Because this is now occurring at regular times, the CPU requirements and output reporting from this QMV is fixed. Each time QMV looks at the latest Kp value it will carry out the same tests. These tests will be a set of sequential operations that each generate yes/no type of information and generates a single fixed record length report. The tests themselves could be simple; is the Kp positive but less than Kp max, since the last Kp value has the Kp changed by ΔKp max, etc. Such tests can be viewed as simply FORTRAN "IF" conditional tests.

The output generated by this QMV would be a fixed size and predictable. Each record would be a set of flags that indicate which tests failed. It would be relatively straightforward to envisage off-line software that could sum up these records to produce daily, weekly, monthly, etc reports, which could be as simple as a single number; a percentage, indicating the accumulated error.

So far, this percentage would not identify the source of the error, just that the Kp data set has errors. However, with this QMV scenario a similar simple QMV would be looking at other output data sets. In this case, if the Kp data sets had problems one would check to see if the merged magnetometer data also had errors. The merged magnetometer data base is a necessary input to the Kp generator (see Figure 2a). Hence, the identification of the source of errors would be based upon a comparison of QMV reports from these two utilities.

The second part of the Kp QMV activity would be to verify that the Kp being generated is equivalent to the Gottingen Kp. As already pointed out, this cannot be done in real-time. Hence, it is off-line activity. This simply requires that the Kp data base at the standard 3 hourly UT intervals is compared with the Gottingen Kp. Such a cross correlation would be scheduled monthly or annually and would involve straightforward off-the-shelf analyses.

This QMV procedure would be developed to evaluate how well a particular utilities output matched the design specification. In the event that the problem had arisen with an input to the utility, a separate QMV procedure would have already identified this "error". Table 3 outlines the advantages of this type of QMV procedure relative to the disadvantages given in Table 2 for the Section 2 QMV. Figure 3 shows how this QMV procedure would be implemented to track the SEC Kp, while Figure 4 shows how the 50th W.W. Kp could be handled. In both cases the QMV can be turned on/off with no impact on the product generating utilities and in no instance can the QMV arbitrarily grow in its need for more CPU resources or output size.

Table 3. Advantages of Independent QMV utilities.

- Requires only a knowledge of the software product, not how the product was generated or its possible failure modes.

- Consists of a sequential set of tests; hence a fixed number of operations, which means no unspecified CPU requirements.

- Output is a fixed record.

- QMV report deals with how well the real time utilities output met the design specification.

- At this level QMV is not carrying out "damage control". It provides the necessary information to readily establish the degree of damage in reports as simple as on number/month if desired.

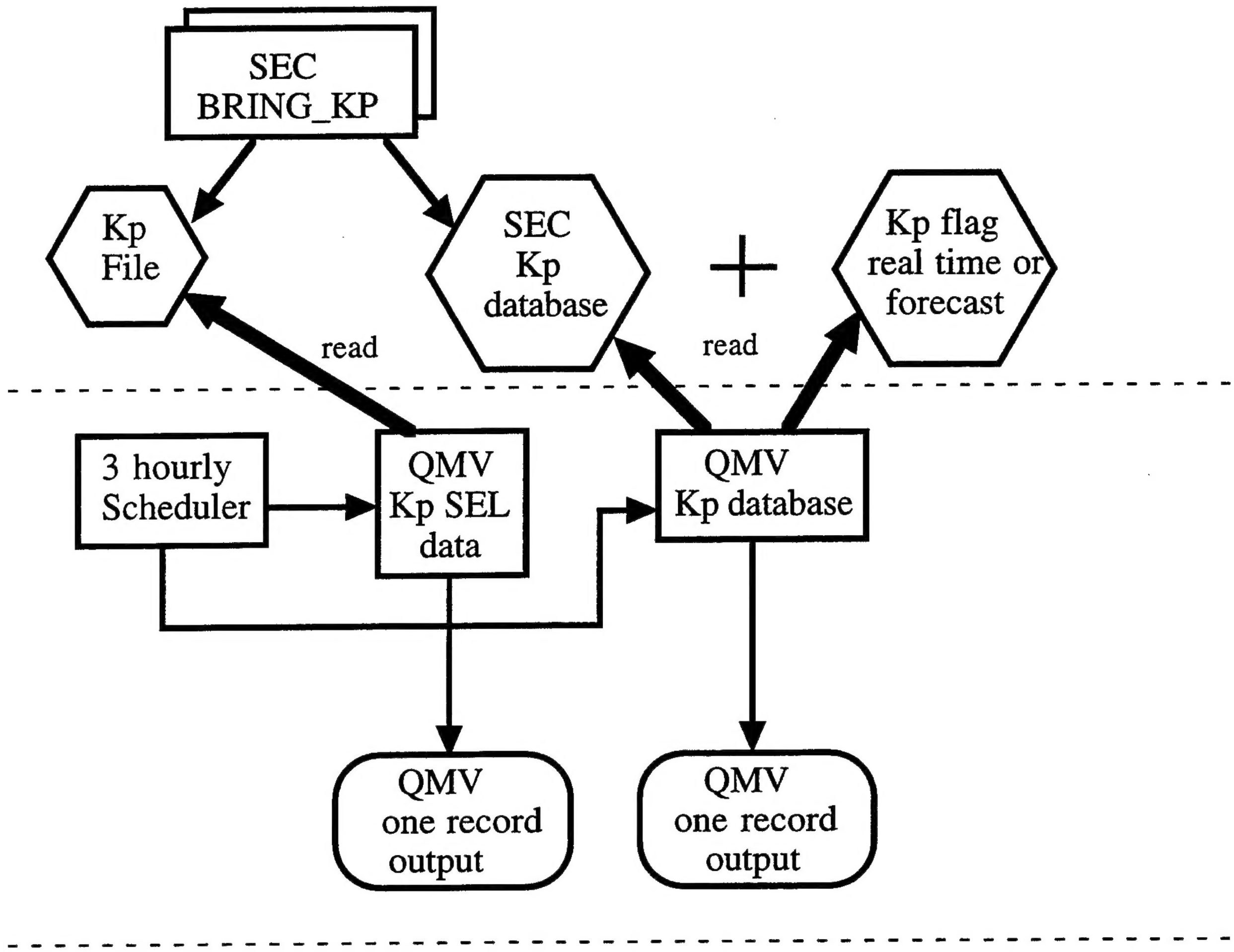

SEC
BRING_KP
Kp
File
SEC
Kp
database
+
Kp flag
real time or
forecast
read
read
3 hourly
Scheduler
QMV
Kp SEL
data
QMV
Kp database
QMV
one record
output
QMV
one record
output

Figure 3

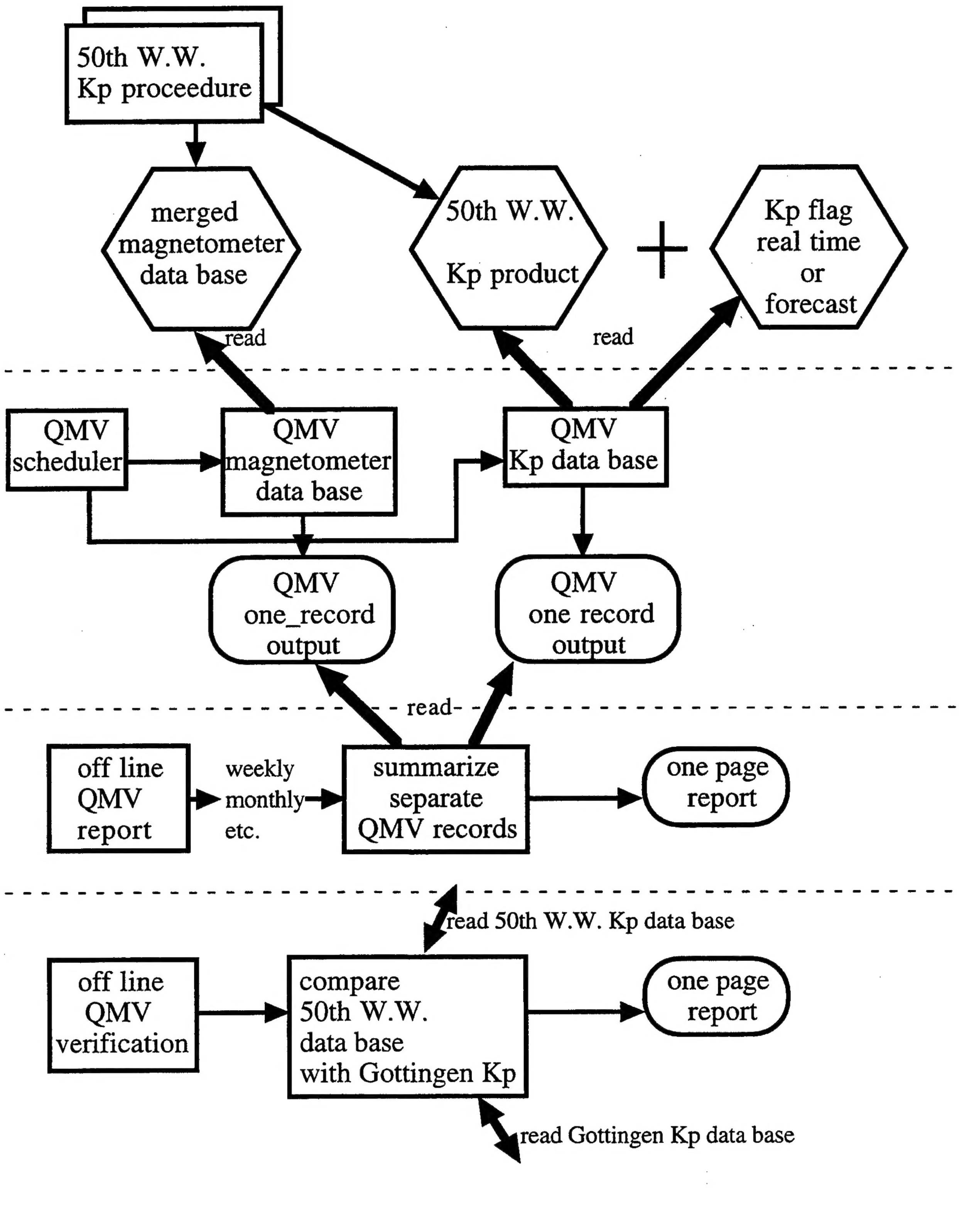

50th W.W.
Kp proceedure
merged magnetometer data base
50th W.W. Kp product
+
Kp flag real time or forecast
read
read
QMV scheduler
QMV magnetometer data base
QMV Kp data base
QMV one_record output
QMV one record output
read-
off line QMV report
weekly monthly etc.
summarize separate QMV records
one page report
read 50th W.W. Kp data base
off line QMV verification
compare 50th W.W. data base with Gottingen Kp
one page report
read Gottingen Kp data base

Figure 4